Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoint

Original Papers

Corrigenda and Addenda

Review

Electronic Monitoring Systems for Hand Hygiene: Systematic Review of Technology

Chaofan Wang[1], BSc, MSc; Weiwei Jiang[1], BSc, MSc; Kangning Yang[1], BSc, MSc; Difeng Yu[1], BSc; Joshua Newn[1], BSc, MSc, PhD; Zhanna Sarsenbayeva[1], BSc, MSc, PhD; Jorge Goncalves[1], BSc, MSc, PhD; Vassilis Kostakos[1], BSc, PhD

School of Computing and Information Systems, The University of Melbourne, Carlton, Australia

Corresponding Author:
Chaofan Wang, BSc, MSc
School of Computing and Information Systems
The University of Melbourne
700 Swanston Street
Carlton, 3053
Australia
Phone: 61 390358966
Email: chaofanw@student.unimelb.edu.au

Abstract

Background: Hand hygiene is one of the most effective ways of preventing health care–associated infections and reducing their transmission. Owing to recent advances in sensing technologies, electronic hand hygiene monitoring systems have been integrated into the daily routines of health care workers to measure their hand hygiene compliance and quality.

Objective: This review aims to summarize the latest technologies adopted in electronic hand hygiene monitoring systems and discuss the capabilities and limitations of these systems.

Methods: A systematic search of PubMed, ACM Digital Library, and IEEE Xplore Digital Library was performed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Studies were initially screened and assessed independently by the 2 authors, and disagreements between them were further summarized and resolved by discussion with the senior author.

Results: In total, 1035 publications were retrieved by the search queries; of the 1035 papers, 89 (8.60%) fulfilled the eligibility criteria and were retained for review. In summary, 73 studies used electronic monitoring systems to monitor hand hygiene compliance, including application-assisted direct observation (5/73, 7%), camera-assisted observation (10/73, 14%), sensor-assisted observation (29/73, 40%), and real-time locating system (32/73, 44%). A total of 21 studies evaluated hand hygiene quality, consisting of compliance with the World Health Organization 6-step hand hygiene techniques (14/21, 67%) and surface coverage or illumination reduction of fluorescent substances (7/21, 33%).

Conclusions: Electronic hand hygiene monitoring systems face issues of accuracy, data integration, privacy and confidentiality, usability, associated costs, and infrastructure improvements. Moreover, this review found that standardized measurement tools to evaluate system performance are lacking; thus, future research is needed to establish standardized metrics to measure system performance differences among electronic hand hygiene monitoring systems. Furthermore, with sensing technologies and algorithms continually advancing, more research is needed on their implementation to improve system performance and address other hand hygiene–related issues.

(J Med Internet Res 2021;23(11):e27880) doi:10.2196/27880

KEYWORDS

hand hygiene; hand hygiene compliance; hand hygiene quality; electronic monitoring systems; systematic review; mobile phone

Introduction

Background

Hand hygiene is one of the most effective ways of reducing the transmission of pathogens that cause health care–associated infections (HAIs) [1-3]. HAIs are infections that people acquire in health care settings [4] and are the most crucial challenge to patient safety in health care [5]. HAIs dramatically increase patients' length of stay, costs, mortality, and morbidity worldwide [6,7]. Moreover, HAIs also impose a heavy financial burden on health care systems. Solely in the United States, the estimated annual costs range from US $28 billion to US $45 billion [8]. The hands of health care workers (HCWs) represent the main pathway of pathogen transmission during health care [2], and Stone et al [9] estimated that at least one-third of HAIs can be prevented by achieving better hand hygiene in health care settings.

In 2009, the World Health Organization (WHO) issued the first *WHO guidelines on hand hygiene in health care* to provide a thorough review of evidence on hand hygiene in health care and specific recommendations to improve practices in health care settings [2]. In the guidelines, the WHO summarizes the five key moments when HCWs should ensure hand hygiene [2], as shown in Figure 1. The guidelines also recommend two standard hand hygiene techniques, *handwash with soap and water* for visibly soiled hands and *hand rub with alcohol-based formulation* for routine decontamination of hands [2], as shown in Figure 2.

Figure 1. The key moments when health care workers should perform hand hygiene. Source: World Health Organization: "My 5 Moments for Hand Hygiene" (with permission) [2].

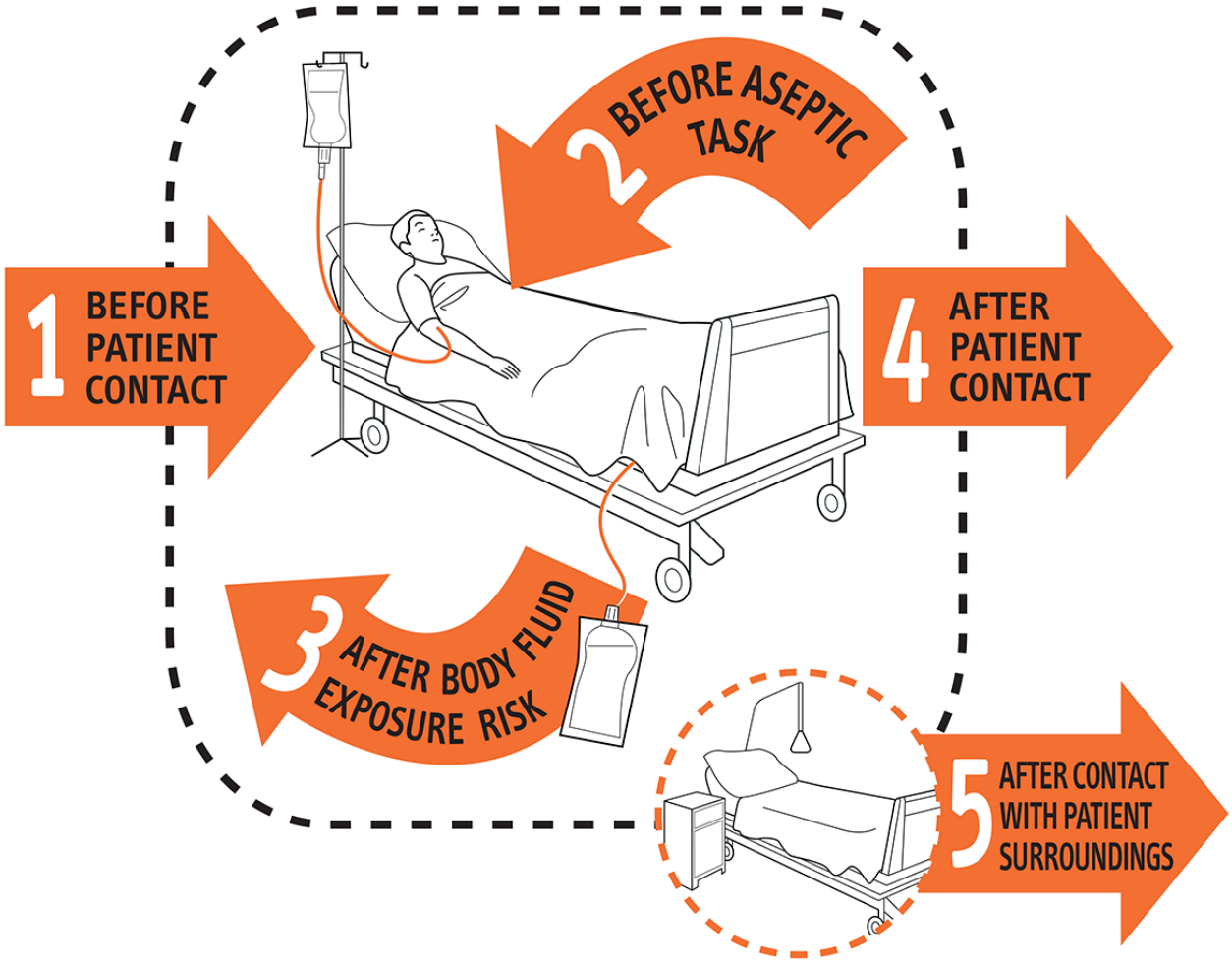


Figure 2. Standard World Health Organization procedures of alcohol-based hand rub and handwash with soap and water. Source: World Health Organization. How to Hand rub?/How to Handwash? (with permission) [10].

However, research has found that hand hygiene compliance is often poor [11,12]. By summarizing 96 empirical studies, Erasmus et al [12] reported that the median compliance rate was only 40% among HCWs. Meanwhile, research also found that hand hygiene quality was unsatisfactory [13-15]. Szilágyi et al [15] reported that only 72% of HCWs could adequately clean all hand surfaces immediately after hand hygiene training. Owing to the importance of hand hygiene, these findings suggest that monitoring hand hygiene practices and providing HCWs with feedback regarding their performance are essential to promote hand hygiene compliance and quality in health care settings [16].

Direct observation by trained auditors is considered the gold standard for monitoring hand hygiene compliance in health care settings [2,17]. Self-reporting by HCWs and the measurement of hand hygiene product consumption are also widely used to monitor hand hygiene compliance [18]. However, Boyce et al [18,19] argued that the disadvantages of direct observation include time and resource consumption, insufficient sample size, lack of standardized observational practices, and the Hawthorne effect. Furthermore, self-reporting is not recommended by experts, as HCWs tend to overestimate their level of compliance, and the measurement of hand hygiene consumption cannot assess the appropriateness of HCWs' hand hygiene timing and quality [18].

To assess hand hygiene quality, previous studies have used direct observation by trained auditors to observe HCWs' compliance with the WHO 6-step hand hygiene technique

[13,14,20]. Another common technique is using UV fluorescent substances to detect the surface coverage of hand hygiene products after hand hygiene [21,22]. Moreover, microbiological tests measure bacteria reduction count to evaluate hand hygiene quality [21,23,24]. However, using direct observation to monitor hand hygiene quality suffers from the same disadvantages as using direct observation to monitor hand hygiene compliance. Visual inspection of fluorescence is restricted to small sample sizes and a lack of standardized observational practices [25]. Furthermore, microbiological tests require time-consuming procedures and often overestimate the reduction of bacteria [21].

Given the above trade-offs, there has been increased interest in developing electronic monitoring systems to serve as an alternative or supplemental monitoring approach [19]. These electronic hand hygiene monitoring systems can be further categorized into electronic hand hygiene compliance monitoring systems and electronic hand hygiene quality monitoring systems.

Although previous reviews have described electronic hand hygiene compliance monitoring systems in detail, this is not the case for electronic hand hygiene quality monitoring systems [19,26,27]. Recent advances in sensor technologies and algorithms have also contributed to the development of new electronic hand hygiene monitoring systems. Furthermore, electronic hand hygiene monitoring systems have limitations that need to be identified and highlighted.

Objectives

This paper aims to (1) review the literature regarding the latest technological developments in electronic hand hygiene systems for monitoring compliance and quality and (2) summarize the limitations and challenges when developing and deploying such systems in health care settings.

Methods

Search Strategy and Selection Criteria

We conducted a bibliographic search of the following web-based databases: PubMed, ACM Digital Library, and IEEE Xplore Digital Library. This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [28] guidelines to reduce the risk of bias and increase its transparency and replicability. This systematic review is not registered on the network, and its review protocol is described below.

We derived the search query using a combination of key terms from previously published literature and expert advice. For the health-related database (PubMed), we specified search terms regarding hand hygiene, technological innovation, and observation to target electronic hand hygiene monitoring systems. For the technological databases (ACM Digital Library and IEEE Xplore Digital Library), we specified terms related to hand hygiene to include relevant technical innovations. The search queries for each database are given in Multimedia Appendix 1. Papers published between January 1, 2000, and June 30, 2020, were included in this study. As older literature is less relevant to today's electronic hand hygiene monitoring systems, we decided to exclude it.

Studies were included if they (1) developed an electronic method or system to monitor hand hygiene compliance or hand hygiene quality, (2) used an existing electronic device or application to support hand hygiene monitoring, or (3) adopted an existing electronic hand hygiene monitoring system and provided sufficient technical details. Meanwhile, studies were excluded if they (1) did not explicitly target electronic hand hygiene monitoring, (2) did not provide adequate technical details (eg, communication protocol and sensor specification), (3) were not published in English, or (4) were not original research papers (eg, abstracts, review papers, and editorials).

To identify the relevant studies, we first imported the search results into a spreadsheet for duplicate removal. Then, the titles were screened based on the selection criteria. If a publication passed the title screening, its abstract was assessed. Finally, the decision for inclusion was made according to the full text of the study. A total of 2 authors, CW and WJ, independently performed the study selection procedure for the retrieved publications. Disagreements between the 2 authors were further summarized and resolved by discussion with the senior author, VK, whenever necessary.

Data Extraction and Data Analysis

To collect information from the included studies in a consistent manner, we created a data extraction table (Multimedia Appendix 2). A total of 2 authors, CW and WJ, independently performed the data extraction procedure, whereas disagreements were resolved by discussion with the senior author, VK.

As we aimed to summarize the different technologies used in electronic hand hygiene monitoring systems, we adopted a narrative approach to synthesize the extracted data. All studies were first grouped by their study aims (monitoring either hand hygiene compliance or quality). After that, the categorized studies were further divided into several categories according to their technical details. Specifically, electronic hand hygiene compliance monitoring systems include (1) application-assisted direct observation, (2) camera-assisted observation, (3) sensor-assisted observation, and (4) real-time locating systems (RTLSs). Meanwhile, electronic hand hygiene quality monitoring systems include (1) measure compliance with the WHO 6-step hand hygiene techniques and (2) detect surface coverage or illumination reduction of fluorescent substances.

Owing to the high level of heterogeneity of the included studies, this study could not provide meta-analyses of the system performance and relevant HCWs' behavior changes. The significant heterogeneity also resulted in missing standardized automation tools to evaluate the risk of bias and assess the certainty for each included study.

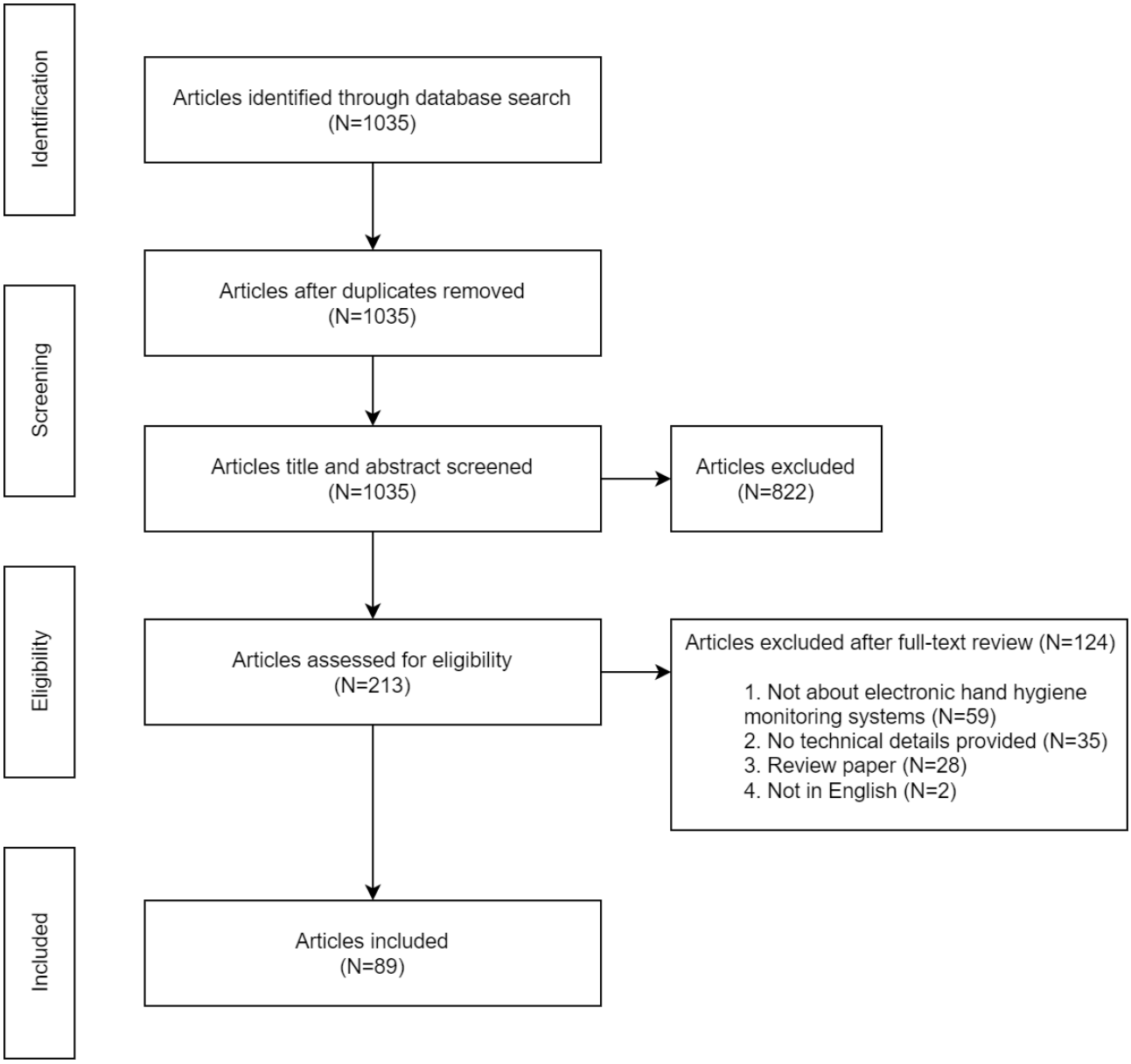
Results

Inclusion of Studies and Study Characteristics

In total, 1035 publications were retrieved by the initial search queries (777/1035, 75.07% from PubMed; 190/1035, 18.36% from the IEEE Xplore Digital Library, and; 68/1035, 6.57% from the ACM Digital Library). None of the retrieved studies were removed based on duplication. After screening the titles and abstracts, 79.42% (822/1035) of studies were excluded for

not meeting the eligibility criteria. Thus, 20.58% (213/1035) of studies were reviewed for the full text. Of these 213 studies, 124 (58.2%) studies were excluded. The main reasons for exclusion were the irrelevance of electronic hand hygiene monitoring systems (59/124, 47.6%) and insufficient technical details (35/124, 28.2%). No study was excluded if they met the inclusion criteria. Therefore, of the 213 studies, 89 (41.8%) fulfilled the eligibility criteria and were retained for review [25,29-116]. Figure 3 shows the process of searching for and selecting the studies included in the review.

Figure 3. Flowchart of the selection process for the systematic review.



All the 89 reviewed studies were published between 2009 and 2020, with 9 (10%) dated in or before 2010 [42,46,56,60,81,83,86,92,113], 38 (43%) dated between 2011 and 2015 [32,33,36,38,39,45,48,49,51,52,54,59,61,63-65,67,68,72,73,78-80,82,84,87-89,93-97,99-101,108,116], and 42 (47%) dated in or after 2016 [25,29-31,34,35,37,40,41,43,44,47,50,53,55,57,58,62,66,69-71,74-77,85,90,91,98,102-107,109-112,114,115]. Regarding the countries where the studies were conducted, 6 countries had ≥5 studies: United States (31/89, 35%) [25,30,31,36,38,39,41,46-48,51,52,59,61,63,68, 69,71,74,76,80-84,92,95,96,101,107,111], Canada (8/89, 9%) [42,72,86-89,98,100], Japan (7/89, 8%) [44,45,55,58,62,114,115], Brazil (6/89, 7%) [32,33,49,56,67,78], Germany (6/89, 7%) [37,40,50,66,108,109], and India (5/89, 6%) [64,65,94,99,102]. The demographic information of participants was provided in only 70% (62/89) of studies. Most studies (50/89, 56%) recruited HCWs from hospitals or clinics [29,30,32-40,44,46-57,59,61,66-68,73-75,77-80,82,84,86,87,89-91,95,96,98-100,103,106], and few (2/89, 2%) studies also involved patients from hospitals [66,94]. The remaining studies recruited the general public (7/89, 8%) [45,60,64,65,109-111] or students (4/89, 4%) [76,104,114,115] from communities or educational settings.

Compliance Monitoring Systems

We identified 73 studies that either implemented or adopted an electronic monitoring system for hand hygiene compliance and grouped them into 4 categories based on their enabling

technology [19,26]: application-assisted direct observation (5/73, 7%), camera-assisted observation (10/73, 14%), sensor-assisted observation (29/73, 40%), and RTLS (32/73, 44%).

Application-Assisted Direct Observation

Approximately 7% (5/73) of studies used applications to assist trained auditors in observing hand hygiene compliance (details are included in Table 1) [29-33]. With these applications, human observers could record their observations using smartphones or tablets. Unlike manual observation with paper forms, application-assisted observation avoids the need for transcription, which could cause delays in analysis, increase the associated cost, and introduce errors [117]. In addition, the prevalence of smartphones and tablets in health care settings makes data collection more unobtrusive and reduces the Hawthorne effect [26]. Both in-house and commercial applications have been used for application-assisted direct observations.

The monitored hand hygiene opportunities may vary in different studies. Most studies followed the instructions given by the WHO *5 moments for hand hygiene* [29,31-33]. Conversely, Sickbert-Bennett et al [30] simplified the observation process to *patient room entry or exit* events (as proxies for moments 1, 4, and 5).

Table 1. Description of application-assisted direct observation studies.

Paper and system description	Required device	System type	System metrics (hand hygiene opportunities)
Kariyawasam et al [29]			
Self-developed application	Android tablet	Research	WHO[a] 5 moments
Magnus et al [32] and Sodré da Costa et al [33]			
iScrub	iOS devices	Commercial	WHO 5 moments
Sickbert-Bennett et al [30]			
iScrub	iOS devices	Commercial	Patient room entry/exit events
SelectSurvey	Web browser	Commercial	Patient room entry/exit events
Wiemken et al [31]			
Google forms	Web browser	Commercial	WHO 5 moments

[a]WHO: World Health Organization.

Camera-Assisted Observation

In contrast with application-assisted direct observation, which solely relies on human auditors, studies with camera-assisted observation could rely on either human auditors [34-40] or algorithms [41-43] for analysis (details included in Table 2). Approximately 30% (3/10) of studies installed cameras inside and outside patient rooms to capture all five hand hygiene moments [34-36]. Researchers manually coded the streaming and recorded videos. Armellino et al [38,39] recruited a remote video auditing company (Arrowsight, Inc) to conduct compliance observations only when HCWs entered or exited the patient room (as proxies for moments 1, 4, and 5). Rather than installing cameras in the environment, Diefenbacher et al [37,40] proposed mounting a camera on the chest of HCWs that aimed at their hands, and researchers further analyzed these first-person view video recordings according to the WHO 5 moments for hand hygiene.

In terms of automated analyses, Zhong et al [41] attached a red green blue (RGB) camera to the chest of HCWs to collect egocentric videos. By feeding RGB images and optical flow images inside a two-stream convolutional neural network, they identified hand hygiene events in HCWs' daily routines [41]. Snoek et al [42] used an RGB camera with a microphone to observe handwash events in older adults with Alzheimer disease. Awwad et al [43] used an RGB-depth camera, Kinect (Microsoft Corporation), to achieve automatic detection of moment 1 (before touching a patient). Depth cameras generate pictures with stereo information, and these pictures have pixels with a value being the distance from the camera or *depth*. Hand hygiene compliance of moment 1 was then estimated by measuring the proximity between the subjects' hands and patient/bed with the presence of hand rub events [43].

Table 2. Description of camera-assisted observation studies.

Paper and system description	Device location	Video type	System type	System metrics (hand hygiene opportunities)
Auditor (human)				
Brotfain et al [34]				
RGB[a] camera	Patient room	Streaming	Research	WHO[b] 5 moments
Sánchez-Carrillo et al [35]				
RGB camera	Patient room	Recorded	Research	WHO 5 moments
Diller et al [36]				
RGB camera (with infrared lens)	Patient room	Recorded	Research	WHO 5 moments
Armellino et al [38,39]				
RGB camera	Sink and sanitizer dispenser	Recorded	Commercial (Arrowsight)	Patient room entry/exit events
Motion sensor	Patient room entrance	Recorded	Commercial (Arrowsight)	Patient room entry/exit events
Diefenbacher et al [37,40]				
RGB camera	HCWs[c] (chest)	Recorded	Research	WHO 5 moments
Auditor (algorithm)				
Zhong et al [41]				
RGB camera	HCWs (chest)	Recorded	Research	Hand hygiene events
Snoek et al [42]				
RGB camera	Sink	Recorded	Research	Hand hygiene events (5 states related to faucet interaction)
Microphone	Sink	Recorded	Research	Hand hygiene events (5 states related to faucet interaction)
Awwad et al [43]				
RGB depth camera	Patient bed	Recorded	Research	Moment 1 (before touching a patient)

[a]RGB: red green blue.

[b]WHO: World Health Organization.

[c]HCW: health care worker.

Sensor-Assisted Observation

Of the 73 studies, 29 (40%) observed hand hygiene compliance using sensors (details are included in Table 3) [32,33,44-70]. These studies were grouped into 3 categories on the basis of sensor type: electronic dispenser, electronic dispenser assisted by other sensors, and inertial measurement unit (IMU) and microphone.

Of these 29 studies, 15 (52%) used electronic dispensers to record the frequency of hand hygiene events and estimate the volume of hand hygiene products dispensed [32,33,44-50,54-57,59,67]. A range of sensors was used to trigger the electronic dispenser counter, including pressure resistors [45,53], magnetic sensors [66], and photosensors [58]. These records were then manually collected by researchers or wirelessly transmitted to the associated servers for further analysis. Compared with direct observation, electronic dispensers can capture hand hygiene events with substantially fewer personnel resources and are unaffected by the Hawthorne effect [19]. However, electronic dispensers cannot detect the hand hygiene opportunities specified by the WHO 5 moments for hand hygiene [19]. Thus, several studies supplemented hand hygiene events with further information to estimate hand hygiene compliance, including outpatient visit records, the expected number of hand hygiene events, ward-specific conversion factors, the number of patients in the unit, nurse visit records, and documented activities [44,48-50,55,57,59].

As electronic dispensers cannot detect hand hygiene opportunities according to the WHO guidelines, other sensors were used to capture these opportunities [51-53,61,62,66,68]. A common practice was to use motion sensors to record patient room entry or exit events (as proxies for moments 1, 4, and 5) [51,61,66]. Here, the dispensers and motion sensors uploaded the time stamp of dispense and room entry/exit events to a server. Once a motion sensor was activated, the server measured the occurrence of a hand hygiene event within a predefined

period and thus estimated the hand hygiene compliance rate. Conversely, Geilleit et al [53] placed a motion sensor around HCWs' working area and pressure plates on patient couches and chairs. Hand hygiene opportunities were defined as the movement of HCWs into a patient zone when the pressure plates were activated. Furthermore, studies used electronic dispensers with other sensors, including IMUs and microphones, to recognize different types of hand hygiene events from HCWs' daily routines [52,62,68].

Of the 29 studies, 7 (24%) used an IMU and microphone to distinguish hand hygiene events from daily activities [58,60,63-65,69,70]. An IMU is an electronic sensor that measures a body's specific force, angular rate, and orientation. Of the 7 studies, 2 (29%) attached an IMU wristband to users' wrists to collect physical signals and utilize these signals to recognize hand hygiene events [63,69]. By using acceleration, gyration, and audio signals from participants' wrists, Wijayasingha et al [70] applied the naive Bayes algorithm to identify both hand hygiene and oral hygiene events from people with developmental disabilities. Instead of placing sensors on users' wrists, 43% (3/7) of studies embedded IMU sensors with or without microphones inside soap bars [60,64,65]. These augmented soap bars were then distributed to low-income households to monitor their soap use associated with hand and body wash. Furthermore, Miyazaki et al [58] attached a microphone to a sink to distinguish hand hygiene events from other daily activities.

Table 3. Description of sensor-assisted observation studies.

Paper and system description	Device location	System type	System metrics (hand hygiene opportunities)
Electronic dispenser			
Arai et al [44]			
Dispenser	Outpatient area	Commercial (Compleo-IO)	Outpatient visit records
Asai et al [45]			
Dispenser	Hospital entrance	Research	Hand hygiene events
Boyce et al [46]			
Dispenser	Patient room and hallway	Commercial (iSIGNOL)	Hand hygiene events
Cohen et al [47]			
Dispenser	Throughout entire facility	Commercial (DebMed GMS)	Hand hygiene events
Conway et al [48]			
Dispenser	Throughout entire facility	N/A[a]	Expected hand hygiene events
Diefenbacher et al [50]			
Dispenser	Patient room	Commercial (Ingo-man We-co)	Hand hygiene events (conversion factor)
Helder et al [54]			
Dispenser	Patient room	Commercial (ComSens NewCompliance)	Hand hygiene events
Kato et al [55]			
Dispenser	Outpatient area	Commercial (CARECOM Co, Ltd)	Outpatient visit records
Morgan et al [59]			
Dispenser	Patient room	N/A	Patients number
De MacEdo et al [49]			
Dispenser	Patient room	Commercial (NXT 1-L model)	Nurse visits (nurse call system)
Marra et al [56,67], Magnus et al [32], and Sodré da Costa et al [33]			
Dispenser	Patient room	Commercial (NXT 1-L model)	Hand hygiene events
Scheithauer et al [57]			
Dispenser	Throughout entire facility	Commercial (Ingo-man We-co)	Documented activities
Electronic dispenser assisted by other sensors			
Ellison et al [51]			
Dispenser	Throughout intensive care units	N/A	Patient room entry/exit events
Motion sensor	Patient room entrance	N/A	Patient room entry/exit events
Sharma et al [61]			
Dispenser	Hallway	Research	Patient room entry/exit events
Motion sensor	Examination room entrance	Research	Patient room entry/exit events
Gaube et al [66]			

Paper and system description	Device location	System type	System metrics (hand hygiene opportunities)
Dispenser	Patient room and hallway	Research	Patient room entry/exit events
Motion sensor	Dispenser	Research	Patient room entry/exit events
Geilleit et al [53]			
Dispenser	Examination room	Research	Patient room entry/exit events
Motion sensor	HCWs'[b] work area	Research	Patient room entry/exit events
Pressure plate	Examination couch, Chair	Research	Patient room entry/exit events
Galluzzi et al [52,68]			
Dispenser	N/A	Research	Hand hygiene events
IMU[c]	HCWs (wristwatch)	Research	Hand hygiene events
Tobita et al [62]			
Dispenser	Sink	Research	Hand hygiene events
Microphone	Sink	Research	Hand hygiene events
IMU and microphone			
Uddin et al [63]			
IMU	HCWs (wristband)	Research	Hand hygiene events
Li et al [69]			
IMU	HCWs (wristband)	Research	Hand hygiene events
Ram et al [60]			
IMU	Soap bar	Research	Hand hygiene events
Wright et al [64] and Zillmer et al [65]			
IMU	Soap bar	Research	Hand hygiene events
Microphone	Soap bar	Research	Hand hygiene events
Wijayasingha et al [70]			
IMU	HCWs (wristwatch)	Research	Hand hygiene events
Microphone	HCWs (wristwatch)	Research	Hand hygiene events
Miyazaki et al [58]			
Microphone	Sink	Research	Hand hygiene events

[a]N/A: not applicable.

[b]HCW: health care worker.

[c]IMU: inertial measurement unit.

Real-time Locating Systems

Of the 73 studies, 32 (44%) studies deployed RTLSs to track hand hygiene compliance (details included in Table 4) [67,71-101]. The RTLS was originally used to identify and track the location of objects or people in real time within a specified area. By sensing dispenser actuation and HCWs' movements, servers from an RTLS can measure HCWs' hand hygiene compliance rates as the ratio of dispenser actuation to the patient room or area entry or exit events (as proxies for moments 1, 4, and 5) [19]. On the basis of the underlying technology, we divided these systems into 6 categories: radio-frequency identification (RFID), infrared, ultrasound, Bluetooth low energy (BLE), IEEE 802.15.4/ZigBee, and Wi-Fi.

Of the 32 studies, 10 (31%) developed or deployed an RFID-based RTLS [71,73,75,76,83,84,90,93,97,99]. RFID uses radio waves to identify and track tags attached to objects. RFID tags can come in a variety of shapes and can be embedded into HCWs' name tags, wristbands, bracelets, and even shoes. When HCWs with RFID tags pass RFID readers, the readers detect the HCWs' tags and then communicate the collected information to a central server. To record HCWs' hand hygiene events, 30% (3/10) of studies placed RFID readers either next to dispensers or embedded RFID readers into dispensers [71,76,93], and these RFID readers were activated by dispensing events. By installing RFID readers next to dispensers and at the entrance of patient rooms or around patient beds, RFID-based RTLS could recognize both hand hygiene events and the entry and exit of

individuals into a patient room or a patient area [73,75,83,84,90,99]. Furthermore, several other sensors were used to assist in the observation of RFID-based RTLS, including motion sensors (recording movements around patient beds), IMUs (recording duration of hand hygiene events), and ethanol sensors (recognizing alcohol-based hand rub events) [90,97,99].

Of the 32 studies, 8 (25%) studies adopted infrared-based RTLSs to monitor hand hygiene compliance [72,74,77,86-89,98]. An infrared transmitter uses infrared light pulses to transmit a unique infrared code to its receiver, and the receiver can then estimate their relative position inside a building. For all 8 studies, infrared transmitters were installed across health care settings and continuously emitted their relative location information (eg, patient room, patient bed, and hallway). In addition, the transmitters were embedded in dispensers and activated for a short period after dispenser actuation. The infrared receivers were carried by HCWs and continuously received location information from the transmitter and counted HCWs' hand hygiene events. In addition, 63% (5/8) of studies used a wearable dispenser to facilitate HCWs' hand hygiene practices. Furthermore, an ethanol sensor was deployed in an infrared-based RTLS to recognize hand rub events rather than relying on a wall-mount dispenser [77].

Of the 32 studies, 2 (6%) studies deployed ultrasound-based RTLSs [79,100]. Similar to other RTLS, ultrasound-based RTLSs comprise transmitters, receivers, and dispensers. Transmitters emit sound in the ultrasonic range, and receivers detect these sounds and thus locate the transmitters. Unlike infrared-based RTLSs, ultrasonic transmitters were typically either placed in health care settings or carried by HCWs, and thus the sound contained either location information or HCWs' identity. Through the collected signals, the receivers could locate the HCWs' real-time location and recognize patient room entry/exit events. When dispensers were used in ultrasound-based RTLSs, transmitters or receivers were also embedded in these dispensers and transmitted dispensing events to the receivers.

Table 4. Description of real-time locating system studies.

Paper and system description	Device location	System type	System metrics (hand hygiene opportunities)
Radio-frequency identification (RFID)			
Decker et al [76]			
RFID tag	HCWs[a] (tag)	Research	Class schedules
RFID reader	Dispenser	Research	Class schedules
Bal et al [71]			
RFID tag	HCWs (tag)	Research	Hand hygiene events
RFID reader	Dispenser and faucet	Research	Hand hygiene events
Dispenser/faucet	Patient room entrance and patient bed	Research	Hand hygiene events
Meydanci et al [93]			
RFID tag	HCWs (wristband)	Research	Hand hygiene events
RFID reader	Dispenser	Research	Hand hygiene events
Dispenser	Patient room and hallway	Research	Hand hygiene events
Boudjema et al [73] and Brouqui et al [75]			
RFID tag	HCWs (shoes)	Commercial (MediHandTrace)	Patient area entry/exit events
RFID reader	Floor (embedded under the dispenser, patient room, and area entrance)	Commercial (MediHandTrace)	Patient area entry/exit events
Dispenser	Patient room and hallway	Commercial (MediHandTrace)	Patient area entry/exit events
Jain et al [83]			
RFID tag	HCWs (tag)	Research	Patient room entry/exit events
RFID reader	Dispenser and patient room entrance	Research	Patient room entry/exit events
Dispenser	Patient room and hallway	Research	Patient room entry/exit events
Johnson et al [84]			
RFID tag	HCWs (tag)	Research	Patient room entry/exit events
RFID reader	Patient room entrance	Research	Patient room entry/exit events
Dispenser	Patient room entrance	Research	Patient room entry/exit events
Radhakrishna et al [99]			
RFID tag	HCWs (tag)	Research	Patient area entry/exit events
RFID reader	Patient trolley (around patient bed)	Research	Patient area entry/exit events
Dispenser	Patient trolley (around patient bed)	Research	Patient area entry/exit events
Motion sensor	Patient trolley (around patient bed)	Research	Patient area entry/exit events
Levin et al [90]			
RFID tag	HCWs (bracelet)	Research	Patient area entry/exit events
RFID reader	Patient bed, Dispenser	Research	Patient area entry/exit events
Dispenser	N/A[b]	Research	Patient area entry/exit events
IMU[c]	HCWs (bracelet)	Research	Patient area entry/exit events
Pleteršek et al [97]			
RFID tag	HCWs (tag)	Research	Hand hygiene events
Ethanol sensor	HCWs (tag), Patient room entrance	Research	Hand hygiene events
Infrared			
Baslyman et al [72]			

Paper and system description	Device location	System type	System metrics (hand hygiene opportunities)
Infrared transmitter	Patient bed, patient room entrance, dispenser, and hallway	Commercial (Ekahau)	Patient area entry/exit events
Infrared receiver	HCWs (tag)	Commercial (Ekahau)	Patient area entry/exit events
Dispenser	Patient room	Commercial (Ekahau)	Patient area entry/exit events
Boyce et al [74]			
Infrared transmitter	Patient bed, dispenser, hallway, and nurse station	Research	Patient area entry/exit events
Infrared receiver	HCWs (tag)	Research	Patient area entry/exit events
Dispenser	N/A	Research	Patient area entry/exit events
Levchenko et al [86-89] and Pong et al [98]			
Infrared transmitter	Individual patient environments, room entrances, shared bathrooms, dirty utility rooms (ceiling), and dispenser	Research	Patient room entry/exit events
Infrared receiver	HCWs (tag)	Research	Patient room entry/exit events
Wall-mount dispenser	N/A	Research	Patient room entry/exit events
Wearable gel dispenser	HCWs	Research	Patient room entry/exit events
Dyson et al [77]			
Infrared transmitter	Patient room and area entrance and sink (ceiling)	N/A	Patient area entry/exit events
Infrared receiver	HCWs (tag)	N/A	Patient area entry/exit events
Ethanol sensor	HCWs (tag)	N/A	Patient area entry/exit events
Ultrasound			
Fisher et al [79]			
Ultrasound transmitter	Patient bed and dispenser	N/A	Patient area entry/exit events
Ultrasound receiver	HCWs (tag)	N/A	Patient area entry/exit events
Dispenser	Patient room	N/A	Patient area entry/exit events
Srigley et al [100]			
Ultrasound transmitter	HCWs (tag)	N/A	Hand hygiene events
Ultrasound receiver	Patient room, hallway, and dispenser	N/A	Hand hygiene events
Dispenser	N/A	N/A	Hand hygiene events
Bluetooth low energy (BLE)			
Karimpour et al [85]			
BLE transmitter	Room	Research	Patient area entry/exit events
BLE receiver	HCWs (smartphone)	Research	Patient area entry/exit events
Misra et al [94]			
BLE transmitter	Patient bed and dispenser	Research	Patient area entry/exit events
BLE receiver	HCWs (smartphone)	Research	Patient area entry/exit events
Dispenser	Patient bed	Research	Patient area entry/exit events
Marques et al [91]			
BLE transmitter	Patient room and area entrance, sink, and dispenser	Research	Patient area entry/exit events
BLE receiver	HCWs (smartphone)	Research	Patient area entry/exit events
Dispenser	Patient room and area entrance	Research	Patient area entry/exit events
IEEE 802.15.4/ZigBee			
Marra et al [67] and Filho et al [78]			

Paper and system description	Device location	System type	System metrics (hand hygiene opportunities)
ZigBee transmitter	HCWs (tag)	Commercial (Infectrack System)	Patient area entry/exit events
ZigBee receiver	Patient bed and dispenser	Commercial (Infectrack System)	Patient area entry/exit events
Dispenser	Patient room	Commercial (Infectrack System)	Patient area entry/exit events
Fries et al [80], Herman et al [81], Hornbeck et al [82], Polgreen et al [92], and Monsalve et al [95,96]			
IEEE 802.15.4 transmitter	Patient bed and dispenser	Research	Patient area entry/exit events
IEEE 802.15.4 receiver	HCWs (tag)	Research	Patient area entry/exit events
Dispenser	Patient room	Research	Patient area entry/exit events
Wi-Fi			
Wan et al [101]			
Wi-Fi transmitter	Room and sink	Research	Hand hygiene events
Wi-Fi receiver	HCWs (tag)	Research	Hand hygiene events
Sink	Room	Research	Hand hygiene events

[a]HCW: health care worker.

[b]N/A: not applicable.

Of the 32 studies, 3 (9%) studies developed RTLSs based on BLE technology [85,91,94]. BLE or Bluetooth is a wireless technology standard used for exchanging data between devices through ultra–high-frequency radio waves. These BLE-based RTLSs also contained transmitters (or beacons), BLE receivers, and dispensers. These transmitters were used as location reference points by placing BLE transmitters in health care settings. BLE receivers brought by HCWs could detect HCWs' real-time location to infer patient room entry/exit events. Unlike the aforementioned RTLSs, BLE receivers could be HCWs' own smartphones instead of carrying additional equipment. To measure hand hygiene events, dispensers triggered the embedded BLE transmitters once they were actuated.

Of the 32 studies, 8 (25%) studies used IEEE 802.15.4 or ZigBee-based RTLSs [67,78,80-82,92,95,96]. IEEE 802.15.4 is a wireless standard capable of low-power, low-cost wireless communication between devices with lower power consumption. ZigBee is a wireless mesh network specification based on the IEEE 802.15.4 standard [118]. Similar to other RTLSs, they comprise transmitters, receivers, and dispensers. Transmitters were either carried by HCWs or placed in a health care environment. Two individual systems were used in the studies, including one commercial system (Infectrack System, i-HealthSys) based on ZigBee and one in-house system based on IEEE 802.15.4. After collecting the relative distance and/or HCWs' identity from transmitters, receivers could identify HCWs' movement when HCWs entered or exited patient areas. The transmitters or receivers were also embedded inside dispensers to recognize hand hygiene events.

The last technology used in RTLSs was Wi-Fi [101]. Wi-Fi is a family of wireless network protocols for building wireless network connections between devices through radio waves. Wi-Fi transmitters were deployed across a room and above a sink, and when HCWs triggered the dispenser next to the sink, the dispenser transmitted the dispensing event to a server through the sink transmitter. The receivers were carried by HCWs, scanned for transmitters in the environment, and periodically uploaded their location to a server.

Quality Monitoring Systems

Of the 89 studies, 21 (24%) studies evaluated hand hygiene quality as performed by HCWs, grouped into 2 categories based on their measurement methods: (1) compliance with the WHO 6-step hand hygiene techniques (14/21, 67%) and (2) surface coverage or illumination reduction of fluorescent substances (7/21, 33%).

Compliance With WHO 6-Step Hand Hygiene Techniques

Of the 21 studies, 14 (67%) studies used a variety of sensors to monitor hand hygiene quality based on compliance with the WHO 6-step hand hygiene techniques (Figure 2). A common practice was to detect the duration of hand hygiene, which is considered a key indicator of quality [13,119]. Furthermore, these systems could recognize HCWs' hand motions as belonging to the individual steps from the WHO 6-step hand hygiene techniques. As such, these systems provided more details regarding HCWs' hand hygiene performance, including missed steps and out-of-order sequences, as noncompliance with all steps of hand hygiene procedures fails to cover all skin surfaces [14,20]. In these studies, sensors were either placed in the environment or attached to HCWs to monitor their hand hygiene performance (details are included in Table 5).

Table 5. Description of studies monitoring compliance with the WHOa 6-step hand hygiene techniques.

Paper and system description	Device location	System type	System metrics (compliance with hand hygiene techniques)
Environmental sensor			
Khan et al [106]			
RGB[b] camera	Sink	Research	Hand hygiene duration
Motion sensor	Sink	Research	Hand hygiene duration
Lacey et al [103]			
RGB camera	Sink	Commercial (SureWash)	An unknown number of individual steps (WHO 6-step hand hygiene technique)
Camilus et al [102]			
Depth camera	Sink	Research	6 individual steps (WHO 6-step hand hygiene technique) and 1 wild hand hygiene technique
Zhong et al [104]			
Depth camera	Sink	Research	9 individual steps (WHO 6-step hand hygiene technique)
Khamis et al [105]			
mmWave radar	Sink	Research	9 individual steps (WHO 6-step hand hygiene technique)
Wearable sensor			
Galluzzi et al [52,68]			
IMU[c]	HCWs[d] (wristwatch)	Research	12 individual steps (WHO 6-step hand hygiene technique), 1 wild hand hygiene technique
Li et al [69]			
IMU	HCWs (wristwatch)	Research	13 individual steps (WHO 6-step hand hygiene technique)
Wijayasingha et al [70]			
IMU	HCWs (wristwatch)	Research	9 individual steps (WHO 6-step hand hygiene technique)
Microphone	HCWs (wristwatch)	Research	9 individual steps (WHO 6-step hand hygiene technique)
Banerjee et al [107]			
IMU	HCWs (armband)	Research	6 individual steps (self-defined hand hygiene technique)
Kutafina et al [108,109]			
IMU	HCWs (armband)	Research	9 individual steps (WHO 6-step hand hygiene technique)
sEMG[e]	HCWs (armband)	Research	9 individual steps (WHO 6-step hand hygiene technique)
Wang et al [110]			
IMU	HCWs (armband)	Research	14 individual steps (WHO 6-step hand hygiene technique)
sEMG	HCWs (armband)	Research	14 individual steps (WHO 6-step hand hygiene technique)
Zhong et al [41]			
RGB camera	HCWs (chest)	Research	7 individual steps (self-defined hand hygiene technique)

[a]WHO: World Health Organization.

[b]RGB: red green blue.

[c]IMU: inertial measurement unit.

[d]HCW: health care worker.

[e]sEMG: surface electromyography.

Of the 14 studies, 5 (36%) studies measured compliance with the WHO 6-step hand hygiene techniques by placing sensors in the environment [102-106]. Khan et al [106] placed an RGB camera and a motion sensor above the sink in operation rooms to monitor HCWs' hand hygiene duration. Lacey et al [103] used a commercial automatic video auditing system (SureWash,

GLANTA Ltd) to monitor HCWs' compliance with the WHO 6-step techniques. Camilus et al [102] and Zhong et al [104] installed an RGB-depth camera (Kinect) above a sink to record hand hygiene events. Hand hygiene videos with stereo information were then analyzed by classifying each frame as an individual step from the 6-step hand hygiene techniques. Instead of using optical sensors, Khanmis et al [105] installed an mmWave sensor above a sink to measure hand hygiene performance. The mmWave is a sensing technology for detecting objects and provides the range, velocity, and angle of these objects. By using the generated frames from mmWave signals, they could classify each frame as one of the nine individual steps in line with the 6-step hand hygiene techniques.

Of the 14 studies, 9 (64%) studies monitored compliance with hand hygiene guidelines by attaching wearable sensors to HCWs [41,52,68-70,107-110]. Of these, the IMU was the most popular sensor and was used in 89% (8/9) of studies with several supplementary sensors. As mentioned above, the IMU can measure a body's specific force, angular rate, and orientation. Approximately 44% (4/9) of studies used the IMU of wristwatches to collect physical signals during hand hygiene events and classified hand motion within a certain time frame as one of the several individual steps of the 6-step hand hygiene techniques [52,68-70]. In addition, microphones have been combined with IMUs to evaluate hand hygiene performance, as the additional audio data could further improve the system accuracy [70]. Owing to hygiene reasons, 44% (4/9) of studies used sensor armbands (Myo armband, North Inc) with IMU to detect HCWs' compliance with hand hygiene techniques [107-110]. Of these 4 studies, 3 (75%) studies used both IMU and surface electromyography (sEMG) sensors from Myo armbands to recognize individual steps in line with 6-step hand hygiene techniques [108-110]. The sEMG sensor is an electrochemical sensor that detects biopotentials using electrodes placed on the skin. In contrast to the aforementioned studies, Zhong et al [41] attached an RGB camera to HCWs' chests. The camera recorded HCWs' hand hygiene practices, and then the collected RGB videos were processed by a deep learning algorithm (two-stream convolutional neural network) to classify hand motions into 7 self-defined hand hygiene steps.

Surface Coverage or Illumination Reduction of Fluorescent Substances

Of the 21 studies, 7 (33%) studies used fluorescent substances to automatically examine hand hygiene quality by computer vision algorithms. However, the means of detecting the quality of handwash and hand rub were distinct. For handwash, participants first applied fluorescent dye on their entire hands and then washed their hands with soap and tap water thoroughly. For hand rub, a hand disinfectant was mixed with a fluorescent dye, and participants used the disinfectant to perform an episode of hand rub. Then, their hands were checked under a UV light lamp and photographed using RGB cameras for further analysis. By comparing the disinfected areas that glowed under UV light and were free from pathogens, Lehotsky et al [120] stated that fluorescent substances could highlight the areas of the hand surface that were adequately disinfected with acceptable accuracy (95% sensitivity and 98% specificity). UV tests have been widely used to assess hand hygiene quality in medical education because of their easy application, low associated costs, and well-visible results [22].

There were two main ways to automatically analyze the collected RGB images: detecting illumination reduction before and after an episode of handwash or measuring the surface coverage of fluorescent substances (details included in Table 6). Approximately 29% (2/7) of studies calculated the illumination difference of fluorescent substances before and after an episode of handwash using Adobe Photoshop (Adobe Inc) and MATLAB (The Math Works, Inc) [25,111]. Hand hygiene quality was then measured by the value of illumination difference, where a bigger difference indicates better hand hygiene performance and vice versa.

Of the 7 studies, 5 (71%) studies analyzed the collected images from both handwash and hand rub by measuring the surface coverage of fluorescent substances [112-116]. For hand rub, the hand rub quality was acceptable if all areas were bright without dark spots, therefore suggesting that all parts of the hand were covered homogeneously with disinfectant [22]. Approximately 40% (2/5) of studies focused on measuring the surface coverage of fluorescent substances after hand rub by applying clustering algorithms [112,113]. For handwash, as fluorescent substances contaminated hands in advance, the handwash quality was measured by the range of cleaned hand areas (dark areas). Approximately 60% (3/5) of studies applied specific threshold values or deep learning algorithms to measure handwash quality [114-116].

Table 6. Description of studies monitoring surface coverage or illumination reduction of fluorescent substances.

Paper and system description	Device location	System type	System metrics (illumination reduction or surface coverage)
Illumination reduction			
Deochand et al [25]			
Fluorescent substance	HCWs[a] (hand)	Research	Illumination reduction (whole hand)
UV lamp	Opaque box	Research	Illumination reduction (whole hand)
RGB[b] camera	Opaque box	Research	Illumination reduction (whole hand)
Pellegrino et al [111]			
Fluorescent substance	HCWs (hand)	Research	Illumination reduction (whole hand)
UV lamp	Dark room	Research	Illumination reduction (whole hand)
RGB camera	Dark room	Research	Illumination reduction (whole hand)
Surface coverage			
Srisomboon et al [112]			
Fluorescent substance	HCWs (hand)	Research	Surface coverage (pixel)
UV lamp	Opaque box	Research	Surface coverage (pixel)
RGB camera	Opaque box	Research	Surface coverage (pixel)
Szilágyi et al [113]			
Fluorescent substance	HCWs (hand)	Research	Surface coverage (pixel)
UV lamp	Opaque box	Research	Surface coverage (pixel)
RGB camera	Opaque box	Research	Surface coverage (pixel)
Yamamoto et al [114,115]			
Fluorescent substance	HCWs (hand)	Research	Surface coverage (segment)
UV lamp	Opaque box	Research	Surface coverage (segment)
RGB camera	Opaque box	Research	Surface coverage (segment)
Naim et al [116]			
Fluorescent substance	HCWs (hand)	Research	Surface coverage (pixel)
UV lamp	Opaque box	Research	Surface coverage (pixel)
RGB camera	Opaque box	Research	Surface coverage (pixel)

[a]HCW: health care worker.

[b]RGB: red green blue.

Discussion

Recently, there has been increased interest in developing electronic monitoring systems to serve as an alternative or supplementary hand hygiene monitoring approach [19]. However, electronic hand hygiene monitoring systems do have limitations. The following sections discuss the limitations related to accuracy, data integration, privacy and confidentiality, potential risks, usability, associated costs, and infrastructure improvements [19,121].

System Accuracy

The system accuracy of electronic hand hygiene monitoring systems is the top concern for HCWs [121,122]. However, systems come with different metrics without standardized measurement tools. System accuracy is also affected by technical issues and geometric constraints.

The metrics often vary substantially in different types of electronic hand hygiene monitoring systems. For electronic hand hygiene compliance monitoring systems, the metrics are based on the number of detectable moments for hand hygiene described by the WHO (Figure 1). A total of 4 different metrics were mentioned in the included studies: (1) hand hygiene events, (2) patient room entry/exit events, (3) patient area entry/exit events, and (4) the WHO 5 moments for hand hygiene. Similarly, the metrics for electronic hand hygiene quality monitoring systems are also disparate. One way to measure HCWs' hand hygiene quality is through detecting their compliance with the WHO 6-step hand hygiene techniques (Figure 2). However, different systems often recognize different sets of individual steps of standardized techniques, which can vary between 6 and 14 individual steps. Detecting illumination reduction or surface coverage of fluorescent substances is another way to measure hand hygiene quality; however, different studies come with different metrics. Several systems can detect

pixel or segment levels of fluorescent areas from the collected RGB images; however, others measure the illumination reduction of the entire hand. Therefore, system results may not accurately reflect HCWs' hand hygiene compliance and quality, and results cannot be compared between different studies without further processing.

Technical issues dramatically affect system accuracy. One of the major concerns is hardware limitations, which result in systems not functioning well under certain situations. For instance, infrared-based RTLS could fail to work if an infrared transmitter or receiver taken by a person is obscured by objects or cloths as the infrared wave cannot penetrate opaque materials [123]. Systems using ethanol sensors to track alcohol-based hand rubs cannot sense HCWs' handwash events [77]. Systems solely relying on motion sensors (ie, without user identity) cannot provide information on who enters or exits patient rooms. Other systems also suffer from reflected signals, signal noise, and interference. Moreover, the algorithms used in these systems may introduce a variety of errors. An example is that machine learning algorithms used to recognize HCWs' compliance with WHO 6-step hand hygiene techniques can generate incorrect classifications [110]. In some extreme cases, these algorithms may not correctly recognize any individual steps and provide an entire sequence of erroneous predictions. Thus, both hardware and algorithm limitations need to be considered when implementing hand hygiene monitoring systems, and effective validation of an electronic hand hygiene monitoring system is required to identify associated technical issues.

System accuracy is also influenced by geometric constraints. To protect patient privacy, studies may attach a curtain in front of cameras [36] or point them toward nonsensitive regions only (handwashing sinks and sanitizer dispensers) [38], which may not allow observation of all hand hygiene opportunities and events and further affect system accuracy. Furthermore, systems based on wearable devices are restricted by device position. For example, recent studies have relied on sensor armbands to detect hand hygiene quality; however, their system accuracy is greatly affected by the actual armband position on the arm [110].

Data Integration

The use of multiple types of sensor data and system records raises new challenges for data integration. Systems use multiple sensors to collect more reliable, accurate, and useful information required for hand hygiene monitoring; however, sensor data fusion comes with problems and issues. One of the most common issues is sensor registration and calibration, as individual sensors have their own local reference frames [124]. Studies applied varying technologies to convert different data from multiple sensors (eg, IMU and sEMG) into one reference frame and starting time, including network time protocols, event-based synchronization methods, and their combination [125,126]. During data fusion and calibration, diverse formats of sensor data could also generate noise and ambiguity, resulting in competitive and conflicting errors, and adding redundancy of sensor data is one of the solutions to increase system reliability [124]. Other issues with multiple sensor data include granularity, timescale, and frequency [124].

Integrating hand hygiene data observed by different systems is another challenge. To increase result accuracy and credibility, studies might use multiple complementary systems to monitor hand hygiene compliance or quality among the same group of HCWs. However, the metric for each observation method was different, and a lack of correlation with their results raised concerns regarding data validity [32]. In addition, different data and result formats raise issues of data integration and require conversion. Moreover, systems could simultaneously observe hand hygiene compliance and quality; however, the means to store and retrieve the records of compliance rate and quality are unclear [41,52,68-70].

Privacy and Confidentiality

Privacy and confidentiality are two other major concerns associated with electronic hand hygiene monitoring systems. Privacy concerns are known to influence HCWs' attitudes toward electronic hand hygiene monitoring systems [19]. Some HCWs perceive these systems as an invasion of their privacy and a pretext for constant surveillance of their daily activities, which makes HCWs distrust these systems and refuse to change their hand hygiene behaviors [121]. Electronic hand hygiene monitoring systems also create special challenges regarding patient privacy [127]. Studies using video cameras to monitor all 5 moments of hand hygiene would require constant video surveillance of patients and patient rooms, resulting in violation of patient privacy [26]. However, limited studies have mentioned patient privacy protection before implementing electronic hand hygiene systems. Moreover, constant surveillance through electronic hand hygiene monitoring systems might raise legal issues, resulting in systems that are unpractical in health care settings, especially when involving cameras and microphones.

The continuous collection of personal data in unprecedented volumes also raises data security concerns [128]. During data collection and storage, users' personal information can be exposed to unauthorized third parties, and the collected data can also be modified or altered through communication protocols (eg, Wi-Fi and Bluetooth) [128]. Furthermore, use scenarios of the collected data are another noticeable concern in hand hygiene monitoring systems for HCWs. Ellingson et al [122] noted that HCWs were worried about the potential use of adherence data for punitive purposes. Thus, an efficient communication mechanism should be established to provide information to HCWs on what data will be collected and stored and how data will be used [121].

Potential Risks

HCWs may face some potential risks caused by electronic hand hygiene monitoring systems. One potential risk is UV-related skin and eye damage caused by UV lamps, which are used to detect HCWs' hand hygiene quality [129,130]. Efficient preventive measures should be placed to protect HCWs' safety and control their daily exposure under a threshold limit of 3.0 mJ/cm^2 [129]. Wearable sensors have gained popularity to assess HCWs' hand hygiene quality, especially wristwatches. However, wearing rings, wristwatches, and bracelets could cause hand contamination [131]; therefore, it is challenging to use wristwatches to monitor hand hygiene procedure compliance, as it can possibly defeat the purpose. Moreover, Ward et al [26]

noted that during the demolition and installation of monitoring systems in health care, the released particulates such as mold or fungus might increase the risk of infection.

Another risk of deploying electronic hand hygiene systems is radio-frequency interference (RFI) with medical devices. RFI, known as a subset of electromagnetic interference, has been reported to cause medical device failure because of interference from various emitters of radio-frequency energy [132]. Badizadegan et al [133] reported that RFI could also result in erroneous laboratory results. Specifically, van der Togt et al [134] noted that RFID might induce potentially hazardous incidents in medical devices. To prevent RFI-associated medical device failures, system designers and device manufacturers should ensure conformance with current RFI standards, and on-site electromagnetic interference tests are required during implementation [132].

Usability

Another challenge for implementing electronic hand hygiene monitoring systems in health care is usability, as the technology may interrupt HCWs' daily workflow to ensure the proper functioning of systems. These usability problems consist of hardware and information delivery. Conway et al [121] summarized hardware-associated usability problems of compliance monitoring systems, including wearable tags (1) as heavy, bulky, and difficult to use; (2) requiring battery power, but batteries are not durable with frequent battery failures; and (3) requiring HCWs to wear them in certain positions. Other usability problems, such as limited sensing range and angles, require HCWs to change their behavior to ensure that systems work properly [77].

Similarly, usability issues also exist when delivering HCWs' hand hygiene performance information. For hand hygiene compliance monitoring, systems use different types of instant prompts (eg, visual reminders, auditory reminders, vibrations, face-to-face feedback, and olfactory stimulus) to remind HCWs regarding missed hand hygiene opportunities; however, these prompts are associated with several usability problems. For example, Dyson et al [77] noted that systems using visual prompts with a red light could cause patient anxiety. Regarding instant prompts for inadequate hand hygiene quality, most systems are designed for medical training purposes, and thus, efficient delivery of instant feedback to HCWs about hand hygiene quality and integrating these systems into their daily routines are still open challenges.

Associated Costs and Infrastructure Improvements

Implementing an electronic hand hygiene monitoring system in health care facilities comes with high costs and infrastructure improvements [19,26,121]. Using electronic systems first requires expenditure on equipment and installation costs, which vary with the selected systems [19,26,121]. Morgan et al [59] estimated that the installation of electronic dispenser–assisted systems in a 15-bed intensive care unit requires a cost between US $30,000 and US $40,000. Another study installed 21 video cameras in the hallways and patient rooms of a 17-bed intensive care unit, costing US $50,000 [38]. For community settings, installing a complete set of electronic hand hygiene monitoring systems is not realistic. Instead of fixing sensors in the environment, studies attached wearable sensors to HCWs or embedded sensors into soap bars to track HCWs' hand hygiene events from their daily routines, which are more scalable and economical.

Except for expenditures on equipment and installation costs, maintenance and personnel costs represent a larger part of system-associated costs. Maintenance costs include system updates, hand rub and soap supplies, an increase in monitored HCWs, and replacement of batteries and defective devices [19]. For in-house systems, technology does not guarantee accurate measurements and requires continuous iteration developments, resulting in maintenance costs and increased personnel needs. Application-assisted direct observation and camera-assisted observation with human auditors are associated with high personnel costs, as these systems require in-house or remote auditors to continually observe hand hygiene opportunities and events.

The installation of electronic hand hygiene systems may disrupt physical infrastructure and require infrastructure improvements. Conway et al [121] noted that infrastructure improvements comprise existing dispenser replacement and fixed hard wiring. As wireless network infrastructure also dramatically affects the system performance, it should be arranged and updated when deploying such systems in health care facilities with outdated network infrastructure.

Performance Feedback

An important but sometimes overlooked aim of deploying electronic hand hygiene monitoring systems in health care settings is to provide educational interventions to HCWs and improve their practices. The intervention methods used in the included studies comprised instant prompts and periodic summaries.

To remind HCWs about missed hand hygiene opportunities, systems may provide instant prompts when noncompliance is detected. Instant prompts comprise visual reminders, auditory reminders, vibrations, face-to-face feedback, olfactory stimuli, and their combinations. To improve HCWs' hand hygiene quality, systems also provide instant prompts when detecting hand hygiene events with inadequate quality. Instant prompts include reminding HCWs about missed steps and disordered sequences of the WHO 6-step hand hygiene techniques and visualizing unclean areas from recorded UV test images. Periodic summaries are also widely adopted to improve HCWs' hand hygiene compliance and quality. Systems deliver periodic summaries to HCWs through reports, dashboards, games, notice boards/monitors, face-to-face feedback, and their combinations.

The included studies also delivered hand hygiene feedback by combining both instant prompts and periodic summaries. For example, Ellison et al [51] adopted auditory reminders as instant prompts and delivered periodic summaries through specific monitor screen savers to remind HCWs of hand hygiene compliance.

Nevertheless, each instant prompt type is associated with specific drawbacks. For visual reminders, Dyson et al [77] noted that red light light-emitting diodes (LEDs) on badges might

cause patient anxiety, so the color of badge LEDs should be adjustable and provide an option to disable the LEDs when necessary. Regarding auditory reminders, Baslyman et al [72] noted that sending audible alerts during the night is not acceptable as most patients are sleeping. Face-to-face feedback is associated with the Hawthorne effect, which causes different hand hygiene behaviors from their daily routines [30]. Using unpleasant odors is also not suitable for most health care facilities as they may cause physical discomfort.

Regarding periodic summaries, designing understandable periodic summaries for HCWs with different educational backgrounds is a challenge [121]. Conway et al [48] noted that HCWs or managers might have difficulty reading and interpreting periodic reports with charts. Efficiently disseminating collected information to HCWs and keeping them informed is challenging as well, as many HCWs have reported never or inconsistently receiving their performance information [48]. Moreover, ensuring that periodic summaries are used to drive hand hygiene improvement instead of punishment is another challenge. Hand hygiene improvement might be short-lived and moderate without HCWs' engagement, constant feedback delivery, detailed action plans, and leadership support [121].

By constantly delivering feedback to HCWs and educating HCWs and medical students on the importance of hand hygiene and the correct procedures, HCWs are likely to improve their hand hygiene techniques and habits. In Multimedia Appendices 3 [53,66,67,77,79,89,98,111], 4 [35,38,39,44,48,51,79,106], and 5 [30,51,88,89], we summarize the performance improvements of HCWs in studies that implemented instant prompts, periodic summaries, or their combinations. However, HCWs have diverse feedback needs. For example, Conway et al [121] and Levchenko et al [89] noted that most HCWs prefer instant prompts rather than periodic summaries, and their compliance rates increased immediately after receiving instant prompts. Nevertheless, Levchenko et al [89] also mentioned that a few HCWs improved their compliance only after they reviewed their individual results.

Implications

Owing to the high level of heterogeneity of the included studies, it is difficult to compare and analyze data across studies. A noticeable difference across the included studies was the variety of system metrics. To generate quantitative analyses, a high degree of standardization is required. Thus, standardized metrics across different hand hygiene monitoring systems need to be established based on system hardware limitations and WHO recommendations. For instance, the number of individual steps of the WHO 6-step hand hygiene techniques can be set to 9 in line with the WHO guideline as steps 3, 6, and 7 (shown in Figure 2) require repeats for both hands.

Given the recent advancements in sensing technologies, hand hygiene monitoring systems can adopt previously unused technology infrastructure or sensors to monitor HCWs' hand hygiene performance. For example, the aforementioned systems require a dedicated device being carried by HCWs to trace their indoor locations. Li et al [135] achieved device-free indoor location tracking by using commodity Wi-Fi, which has been installed in most health care facilities. Conversely, hand hygiene monitoring systems can apply new algorithms to improve their system accuracy. For example, previous studies adopted a hidden Markov model to classify the individual steps of 6-step techniques or smooth classification results, which assumes that HCWs will perform hand hygiene procedures according to predefined orders. However, once this assumption is relaxed, the performance of these systems dramatically drops [69]. Instead, classification results smoothed by change point detection algorithms (eg, E.Divisive [136]) might ease the performance decrease.

Hand hygiene monitoring systems and collected data can also be used to solve other hand hygiene–related issues. For example, systems detecting surface coverage of fluorescent substances could be considered as an alternative method to validate the efficacy of newly proposed hand hygiene techniques instead of microbiological tests, as fluorescent substances could highlight the hand surface areas that are adequately disinfected with acceptable accuracy [120]. Similarly, studies have used hand hygiene behavior data to monitor participants' levels of dementia, Alzheimer disease, and obsessive-compulsive disorder [137,138]. Furthermore, hand hygiene compliance history has been used to simulate the transmission of HAIs in health care settings [139].

Limitations

This study has several limitations. Some relevant studies may have been missed because of the keywords and databases chosen for the search query. Furthermore, some relevant studies may not have been included if they were not published in English, were outside the specified time frame, or did not provide adequate technical information.

Specifically, we included all types of studies regardless of their maturity, as it helps summarize the latest technological developments in electronic hand hygiene monitoring systems. However, early-stage or preliminary studies or methodology studies may present incomplete data or a lack of results. Owing to the heterogeneity of the studies and sparse metrics, we could not conduct a meta-analysis for the study population, system accuracy, and intervention effectiveness. In addition, because of the significant heterogeneity, we could not evaluate the risk of bias for each study using standardized automation tools and assess the certainty of the included studies.

This review describes different technologies for hand hygiene monitoring. Nevertheless, since we adopted the narrative approach to synthesize the outcomes rather than a meta-analysis, we did not assess the risk of bias because of missing results.

Conclusions

Our review provides an overview of the latest technological developments in electronic hand hygiene monitoring systems that measure compliance or quality. Systems utilize application-assisted direct observation, camera-assisted observation, sensor-assisted observation, and RTLS to monitor HCWs' compliance rates. For quality monitoring, systems either measure compliance with the WHO 6-step hand hygiene techniques or detect surface coverage or illumination reduction of fluorescent substances. Despite the technologies used in these

systems, we identify system-associated issues and challenges, including system accuracy, data integration, privacy and confidentiality, potential risks, usability, and associated costs and infrastructure improvements. Owing to the narrative approach adopted in these studies, more research is required to establish standardized metrics to measure system performance differences among electronic hand hygiene monitoring systems. With sensing technologies and algorithms continually advancing, more research is needed on their implementation to improve system performance and address other hand hygiene–related issues.

Acknowledgments

This work was partially funded by the Australian Research Council Discovery Project DP190102627 and the National Health and Medical Research Council grants 1170937 and 2004316. CW was supported by a PhD scholarship provided by the Australian Commonwealth Government Research Training Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for PubMed, IEEE Xplore Digital Library, and ACM Digital Library.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i11e27880_app1.xlsx]

Multimedia Appendix 2
Data extraction form.
[XLSX File (Microsoft Excel File), 11 KB - jmir_v23i11e27880_app2.xlsx]

Multimedia Appendix 3
Hand hygiene improvements of studies with instant prompts.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i11e27880_app3.xlsx]

Multimedia Appendix 4
Hand hygiene improvements of studies with periodic summaries.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i11e27880_app4.xlsx]

Multimedia Appendix 5
Hand hygiene improvements of studies with instant prompts and periodic summaries.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i11e27880_app5.xlsx]

References

1. Allegranzi B, Pittet D. Role of hand hygiene in healthcare-associated infection prevention. J Hosp Infect 2009 Dec;73(4):305-315. [doi: 10.1016/j.jhin.2009.04.019] [Medline: 19720430]
2. World Health Organization. WHO guidelines on hand hygiene in health care. Geneva: World Health Organization; 2009:1-270.
3. Pittet D. Improving adherence to hand hygiene practice: a multidisciplinary approach. Emerg Infect Dis 2001 Apr;7(2):234-240 [FREE Full text] [doi: 10.3201/eid0702.010217] [Medline: 11294714]
4. Haque M, Sartelli M, McKimm J, Bakar M. Health care-associated infections - an overview. Infect Drug Resist 2018;11:2321-2333 [FREE Full text] [doi: 10.2147/IDR.S177247] [Medline: 30532565]
5. Institute of Medicine, Committee on Quality of Health Care in America. In: Donaldson MS, Corrigan JM, Kohn LT, editors. To Err Is Human: Building a Safer Health System. Washington (DC): National Academies Press (US); 2000.
6. Klevens RM, Edwards JR, Richards CL, Horan TC, Gaynes RP, Pollock DA, et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. Public Health Rep 2007 Aug 02;122(2):160-166 [FREE Full text] [doi: 10.1177/003335490712200205] [Medline: 17357358]
7. De Angelis G, Murthy A, Beyersmann J, Harbarth S. Estimating the impact of healthcare-associated infections on length of stay and costs. Clin Microbiol Infect 2010 Dec;16(12):1729-1735 [FREE Full text] [doi: 10.1111/j.1469-0691.2010.03332.x] [Medline: 20673257]
8. Graves N, Halton K, Paterson D, Whitby M. Economic rationale for infection control in Australian hospitals. Healthc Infect 2009 Sep;14(3):81-88. [doi: 10.1071/hi09010]

9. Stone PW, Hasan S, Quiros D, Larson EL. Effect of guideline implementation on costs of hand hygiene. Nurs Econ 2007;25(5):279-284 [FREE Full text] [Medline: 18080624]
10. How to handrub? / how to handwash? World Health Organization. 2009. URL: https://www.who.int/gpsc/5may/How_To_HandWash_Poster.pdf [accessed 2021-10-21]
11. Pittet D, Simon A, Hugonnet S, Pessoa-Silva CL, Sauvan V, Perneger TV. Hand hygiene among physicians: performance, beliefs, and perceptions. Ann Intern Med 2004 Jul 06;141(1):1-8. [doi: 10.7326/0003-4819-141-1-200407060-00008] [Medline: 15238364]
12. Erasmus V, Daha TJ, Brug H, Richardus JH, Behrendt MD, Vos MC, et al. Systematic review of studies on compliance with hand hygiene guidelines in hospital care. Infect Control Hosp Epidemiol 2010 Mar 02;31(3):283-294. [doi: 10.1086/650451] [Medline: 20088678]
13. Korhonen A, Ojanperä H, Puhto T, Järvinen R, Kejonen P, Holopainen A. Adherence to hand hygiene guidelines - significance of measuring fidelity. J Clin Nurs 2015 Nov 01;24(21-22):3197-3205. [doi: 10.1111/jocn.12969] [Medline: 26331437]
14. Arias AV, Garcell HG, Ochoa YR, Arias KF, Miranda FR. Assessment of hand hygiene techniques using the World Health Organization's six steps. J Infect Public Health 2016 May;9(3):366-369 [FREE Full text] [doi: 10.1016/j.jiph.2015.11.006] [Medline: 26707705]
15. Szilágyi L, Haidegger T, Lehotsky A, Nagy M, Csonka E, Sun X, et al. A large-scale assessment of hand hygiene quality and the effectiveness of the "WHO 6-steps". BMC Infect Dis 2013 May 30;13(1):249 [FREE Full text] [doi: 10.1186/1471-2334-13-249] [Medline: 23718728]
16. Boyce JM. Measuring healthcare worker hand hygiene activity: current practices and emerging technologies. Infect Control Hosp Epidemiol 2011 Oct 02;32(10):1016-1028. [doi: 10.1086/662015] [Medline: 21931253]
17. Stewardson A, Sax H, Longet-Di Pietro S, Pittet D. Impact of observation and analysis methodology when reporting hand hygiene data. J Hosp Infect 2011 Apr;77(4):358-359. [doi: 10.1016/j.jhin.2010.12.008] [Medline: 21316121]
18. Boyce JM. Hand hygiene compliance monitoring: current perspectives from the USA. J Hosp Infect 2008 Oct;70:2-7. [doi: 10.1016/s0195-6701(08)60003-1]
19. Boyce JM. Electronic monitoring in combination with direct observation as a means to significantly improve hand hygiene compliance. Am J Infect Control 2017 May 01;45(5):528-535. [doi: 10.1016/j.ajic.2016.11.029] [Medline: 28456322]
20. Tschudin-Sutter S, Sepulcri D, Dangel M, Schuhmacher H, Widmer AF. Compliance with the World Health Organization hand hygiene technique: a prospective observational study. Infect Control Hosp Epidemiol 2015 Apr 20;36(4):482-483. [doi: 10.1017/ice.2014.82] [Medline: 25782907]
21. Reilly JS, Price L, Lang S, Robertson C, Cheater F, Skinner K, et al. A pragmatic randomized controlled trial of 6-step vs 3-step hand hygiene technique in acute hospital care in the United Kingdom. Infect Control Hosp Epidemiol 2016 Jun 07;37(6):661-666. [doi: 10.1017/ice.2016.51] [Medline: 27050843]
22. Vanyolos E, Peto K, Viszlai A, Miko I, Furka I, Nemeth N, et al. Usage of ultraviolet test method for monitoring the efficacy of surgical hand rub technique among medical students. J Surg Educ 2015 May;72(3):530-535. [doi: 10.1016/j.jsurg.2014.12.002] [Medline: 25656633]
23. Kampf G, Ostermeyer C, Heeg P, Paulson D. Evaluation of two methods of determining the efficacies of two alcohol-based hand rubs for surgical hand antisepsis. Appl Environ Microbiol 2006 Jun;72(6):3856-3861. [doi: 10.1128/aem.02746-05]
24. Michaud RN, McGrath MB, Goss WA. Application of a gloved-hand model for multiparameter measurements of skin-degerming activity. J Clin Microbiol 1976 Apr;3(4):406-413 [FREE Full text] [doi: 10.1128/jcm.3.4.406-413.1976] [Medline: 1262452]
25. Deochand N, Deochand ME. Brief report on hand-hygiene monitoring systems: a pilot study of a computer-assisted image analysis technique. J Environ Health 2016 Jun;78(10):14-20. [Medline: 27348978]
26. Ward MA, Schweizer ML, Polgreen PM, Gupta K, Reisinger HS, Perencevich EN. Automated and electronically assisted hand hygiene monitoring systems: a systematic review. Am J Infect Control 2014 May;42(5):472-478. [doi: 10.1016/j.ajic.2014.01.002] [Medline: 24773785]
27. Marra AR, Edmond MB. Hand hygiene: state-of-the-art review with emphasis on new technologies and mechanisms of surveillance. Curr Infect Dis Rep 2012 Dec 28;14(6):585-591. [doi: 10.1007/s11908-012-0288-y] [Medline: 22922997]
28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
29. Kariyawasam N, Wong M, Turner P. User centred design and nosocomials in surgical ICUs: a mobile application for peer monitoring and training in hand hygiene. Stud Health Technol Inform 2017;245:743-747. [Medline: 29295197]
30. Sickbert-Bennett EE, DiBiase LM, Willis TM, Wolak ES, Weber DJ, Rutala WA. Reducing health care-associated infections by implementing a novel all hands on deck approach for hand hygiene compliance. Am J Infect Control 2016 May 02;44(5 Suppl):13-16. [doi: 10.1016/j.ajic.2015.11.016] [Medline: 27131129]
31. Wiemken TL, Furmanek SP, Mattingly WA, Haas J, Ramirez JA, Carrico RM. Googling your hand hygiene data: using Google Forms, Google Sheets, and R to collect and automate analysis of hand hygiene compliance monitoring. Am J Infect Control 2018 Jun;46(6):617-619. [doi: 10.1016/j.ajic.2018.01.010] [Medline: 29496338]

32. Magnus TP, Marra AR, Camargo TZ, Victor ED, da Costa LS, Cardoso VJ, et al. Measuring hand hygiene compliance rates in different special care settings: a comparative study of methodologies. Int J Infect Dis 2015 Apr;33:205-208 [FREE Full text] [doi: 10.1016/j.ijid.2015.02.016] [Medline: 25722282]
33. da Costa LS, Neves VM, Marra AR, Camargo TZ, dos Santos Cardoso MF, da Silva Victor E, et al. Measuring hand hygiene compliance in a hematology-oncology unit: a comparative study of methodologies. Am J Infect Control 2013 Nov;41(11):997-1000. [doi: 10.1016/j.ajic.2013.03.301] [Medline: 23769833]
34. Brotfain E, Livshiz-Riven I, Gushansky A, Erblat A, Koyfman L, Ziv T, et al. Monitoring the hand hygiene compliance of health care workers in a general intensive care unit: use of continuous closed circle television versus overt observation. Am J Infect Control 2017 Aug 01;45(8):849-854. [doi: 10.1016/j.ajic.2017.03.015] [Medline: 28479008]
35. Sánchez-Carrillo LA, Rodríguez-López JM, Galarza-Delgado D, Baena-Trejo L, Padilla-Orozco M, Mendoza-Flores L, et al. Enhancement of hand hygiene compliance among health care workers from a hemodialysis unit using video-monitoring feedback. Am J Infect Control 2016 Aug 01;44(8):868-872. [doi: 10.1016/j.ajic.2016.01.040] [Medline: 27068027]
36. Diller T, Kelly JW, Blackhurst D, Steed C, Boeker S, McElveen DC. Estimation of hand hygiene opportunities on an adult medical ward using 24-hour camera surveillance: validation of the HOW2 Benchmark Study. Am J Infect Control 2014 Jun;42(6):602-607. [doi: 10.1016/j.ajic.2014.02.020] [Medline: 24837110]
37. Diefenbacher S, Pfattheicher S, Keller J. On the role of habit in self-reported and observed hand hygiene behavior. Appl Psychol Health Well Being 2020 Mar 28;12(1):125-143. [doi: 10.1111/aphw.12176] [Medline: 31353823]
38. Armellino D, Hussain E, Schilling ME, Senicola W, Eichorn A, Dlugacz Y, et al. Using high-technology to enforce low-technology safety measures: the use of third-party remote video auditing and real-time feedback in healthcare. Clin Infect Dis 2012 Jan 01;54(1):1-7 [FREE Full text] [doi: 10.1093/cid/cir773] [Medline: 22109950]
39. Armellino D, Trivedi M, Law I, Singh N, Schilling ME, Hussain E, et al. Replicating changes in hand hygiene in a surgical intensive care unit with remote video auditing and feedback. Am J Infect Control 2013 Oct;41(10):925-927. [doi: 10.1016/j.ajic.2012.12.011] [Medline: 23489740]
40. Diefenbacher S, Sassenrath C, Tatzel J, Keller J. Evaluating healthcare workers' hand hygiene performance using first-person view video observation in a standardized patient-care scenario. Am J Infect Control 2020 May;48(5):496-502. [doi: 10.1016/j.ajic.2019.11.032] [Medline: 32334725]
41. Zhong C, Reibman AR, Cordoba HM, Deering AJ. Hand-hygiene activity recognition in egocentric video. In: Proceedings of the IEEE 21st International Workshop on Multimedia Signal Processing (MMSP). 2019 Presented at: IEEE 21st International Workshop on Multimedia Signal Processing (MMSP); Sept. 27-29, 2019; Kuala Lumpur, Malaysia. [doi: 10.1109/mmsp.2019.8901753]
42. Snoek J, Taati B, Eskin Y, Mihailidis A. Automatic segmentation of video to aid the study of faucet usability for older adults. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. 2010 Presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops; June 13-18, 2010; San Francisco, CA, USA. [doi: 10.1109/cvprw.2010.5543266]
43. Awwad S, Tarvade S, Piccardi M, Gattas DJ. The use of privacy-protected computer vision to measure the quality of healthcare worker hand hygiene. Int J Qual Health Care 2019 Feb 01;31(1):36-42. [doi: 10.1093/intqhc/mzy099] [Medline: 29767747]
44. Arai A, Tanabe M, Nakamura A, Yamasaki D, Muraki Y, Kaneko T, et al. Utility of electronic hand hygiene counting devices for measuring physicians' hand hygiene adherence applied to outpatient settings. Am J Infect Control 2016 Dec 01;44(12):1481-1485 [FREE Full text] [doi: 10.1016/j.ajic.2016.08.002] [Medline: 27665030]
45. Asai T, Kanazawa A, Hayashi H, Minazuki A. Development of a system to raise awareness of hand hygiene in various environments. In: Proceedings of the International Conference on Signal-Image Technology & Internet-Based Systems. 2013 Presented at: International Conference on Signal-Image Technology & Internet-Based Systems; Dec. 2-5, 2013; Kyoto, Japan. [doi: 10.1109/sitis.2013.150]
46. Boyce JM, Cooper T, Dolan MJ. Evaluation of an electronic device for real-time measurement of alcohol-based hand rub use. Infect Control Hosp Epidemiol 2009 Nov 02;30(11):1090-1095. [doi: 10.1086/644756] [Medline: 19775238]
47. Cohen B, Murray M, Jia H, Jackson O, Saiman L, Neu N, et al. Is hand hygiene frequency associated with the onset of outbreaks in pediatric long-term care? Am J Infect Control 2016 Dec 01;44(12):1492-1494 [FREE Full text] [doi: 10.1016/j.ajic.2016.06.022] [Medline: 27566873]
48. Conway LJ, Riley L, Saiman L, Cohen B, Alper P, Larson EL. Implementation and impact of an automated group monitoring and feedback system to promote hand hygiene among health care personnel. Jt Comm J Qual Patient Saf 2014 Sep;40(9):408-417. [doi: 10.1016/s1553-7250(14)40053-9]
49. Macedo RC, Jacob EM, Silva VP, Santana EA, Souza AF, Gonçalves P, et al. Positive deviance: using a nurse call system to evaluate hand hygiene practices. Am J Infect Control 2012 Dec;40(10):946-950. [doi: 10.1016/j.ajic.2011.11.015] [Medline: 22418606]
50. Diefenbacher S, Fliss P, Tatzel J, Wenk J, Keller J. A quasi-randomized controlled before-after study using performance feedback and goal setting as elements of hand hygiene promotion. J Hosp Infect 2019 Apr;101(4):399-407. [doi: 10.1016/j.jhin.2019.02.001] [Medline: 30738912]

51. Ellison RT, Barysauskas CM, Rundensteiner EA, Wang D, Barton B. A prospective controlled trial of an electronic hand hygiene reminder system. Open Forum Infect Dis 2015 Dec;2(4):ofv121 [FREE Full text] [doi: 10.1093/ofid/ofv121] [Medline: 26430698]
52. Galluzzi V, Herman T, Shumaker DJ, Macinga DR, Arbogast JW, Segre EM, et al. Electronic recognition of hand hygiene technique and duration. Infect Control Hosp Epidemiol 2014 Oct 10;35(10):1298-1300. [doi: 10.1086/678059] [Medline: 25203186]
53. Geilleit R, Hen ZQ, Chong CY, Loh AP, Pang NL, Peterson GM, et al. Feasibility of a real-time hand hygiene notification machine learning system in outpatient clinics. J Hosp Infect 2018 Oct;100(2):183-189. [doi: 10.1016/j.jhin.2018.04.004] [Medline: 29649558]
54. Helder OK, van Goudoever JB, Hop WC, Brug J, Kornelisse RF. Hand disinfection in a neonatal intensive care unit: continuous electronic monitoring over a one-year period. BMC Infect Dis 2012 Oct 08;12(1):248 [FREE Full text] [doi: 10.1186/1471-2334-12-248] [Medline: 23043639]
55. Kato H, Takeda R, Ideno Y, Suzuki T, Sano K, Nakamura K. Physicians' compliance for hand hygiene in medical outpatient clinics: automated hand-hygiene monitoring with touch sensor and wireless internet. Am J Infect Control 2021 Jan;49(1):50-54. [doi: 10.1016/j.ajic.2020.05.037] [Medline: 32512079]
56. Marra AR, Moura DF, Paes T, dos Santos OF, Edmond MB. Measuring rates of hand hygiene adherence in the intensive care setting: a comparative study of direct observation, product usage, and electronic counting devices. Infect Control Hosp Epidemiol 2010 Aug 02;31(8):796-801. [doi: 10.1086/653999] [Medline: 20565261]
57. Scheithauer S, Batzer B, Dangel M, Passweg J, Widmer A. Workload even affects hand hygiene in a highly trained and well-staffed setting: a prospective 365/7/24 observational study. J Hosp Infect 2017 Sep;97(1):11-16. [doi: 10.1016/j.jhin.2017.02.013] [Medline: 28389091]
58. Miyazaki K, Tobita H. SinkAmp: interactive sink to detect living habits for healthcare and quality of life. In: Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion. 2018 Presented at: Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion; March 7 - 11, 2018; Tokyo Japan p. 1-2. [doi: 10.1145/3180308.3180339]
59. Morgan DJ, Pineles L, Shardell M, Young A, Ellingson K, Jernigan JA, et al. Automated hand hygiene count devices may better measure compliance than human observation. Am J Infect Control 2012 Dec;40(10):955-959. [doi: 10.1016/j.ajic.2012.01.026] [Medline: 22633134]
60. Ram PK, Halder AK, Granger SP, Jones T, Hall P, Hitchcock D, et al. Is structured observation a valid technique to measure handwashing behavior? Use of acceleration sensors embedded in soap to assess reactivity to structured observation. Am J Trop Med Hyg 2010 Nov;83(5):1070-1076 [FREE Full text] [doi: 10.4269/ajtmh.2010.09-0763] [Medline: 21036840]
61. Sharma D, Thomas GW, Foster ED, Iacovelli J, Lea KM, Streit JA, et al. The precision of human-generated hand-hygiene observations: a comparison of human observation with an automated monitoring system. Infect Control Hosp Epidemiol 2012 Dec 02;33(12):1259-1261 [FREE Full text] [doi: 10.1086/668426] [Medline: 23143367]
62. Tobita H, Matsumoto M, Maida Y, Eguchi Y, Terada K. SinkNet: Interactive sink to detect living habits for healthcare and quality of life using private networks. In: Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). 2018 Presented at: IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Dec. 10-12, 2018; Taichung, Taiwan. [doi: 10.1109/aivr.2018.00020]
63. Uddin M, Salem A, Nam I, Nadeem T. Wearable sensing framework for human activity monitoring. In: Proceedings of the 2015 workshop on Wearable Systems and Applications`. 2015 Presented at: MobiSys'15: The 13th Annual International Conference on Mobile Systems, Applications, and Services; May 18, 2015; Florence Italy p. 21-26. [doi: 10.1145/2753509.2753513]
64. Wright RL, Zillmer R, Biran A, Hall P, Sidibe M. Use of electronic loggers to measure changes in the rates of hand washing with soap in low-income urban households in India. PLoS One 2015 Jun 23;10(6):e0131187 [FREE Full text] [doi: 10.1371/journal.pone.0131187] [Medline: 26101886]
65. Zillmer R, Wright R, Bates S, Mahers I. A robust device for large-scale monitoring of bar soap usage in free-living conditions. Pers Ubiquit Comput 2014 Feb 1;18(8):2057-2064. [doi: 10.1007/s00779-014-0760-9]
66. Gaube S, Tsivrikos D, Dollinger D, Lermer E. How a smiley protects health: a pilot intervention to improve hand hygiene in hospitals by activating injunctive norms through emoticons. PLoS One 2018 May 21;13(5):e0197465 [FREE Full text] [doi: 10.1371/journal.pone.0197465] [Medline: 29782516]
67. Marra AR, Camargo TZ, Magnus TP, Blaya RP, Dos Santos G, Guastelli LR, et al. The use of real-time feedback via wireless technology to improve hand hygiene compliance. Am J Infect Control 2014 Jun;42(6):608-611. [doi: 10.1016/j.ajic.2014.02.006] [Medline: 24725515]
68. Galluzzi V, Herman T, Polgreen P. Hand hygiene duration and technique recognition using wrist-worn sensors. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks. 2015 Presented at: IPSN '15: The 14th International Conference on Information Processing in Sensor Networks; April 13 - 16, 2015; Seattle Washington p. 106-117. [doi: 10.1145/2737095.2737106]
69. Li H, Chawla S, Li R, Jain S, Abowd G, Starner T, et al. Wristwash: towards automatic handwashing assessment using a wrist-worn device. In: Proceedings of the 2018 ACM International Symposium on Wearable Computers. 2018 Presented

at: UbiComp '18: The 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing; October 8 - 12, 2018; Singapore p. 132-139. [doi: 10.1145/3267242.3267247]
70. Wijayasingha L, Lo B. A wearable sensing framework for improving personal and oral hygiene for people with developmental disabilities. In: Proceedings of the IEEE Wireless Health (WH). 2016 Presented at: IEEE Wireless Health (WH); Oct. 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764550]
71. Bal M, Abrishambaf R. A system for monitoring hand hygiene compliance based-on Internet-of-Things. In: Proceedings of the IEEE International Conference on Industrial Technology (ICIT). 2017 Presented at: IEEE International Conference on Industrial Technology (ICIT); March 22-25, 2017; Toronto, ON, Canada. [doi: 10.1109/icit.2017.7915560]
72. Baslyman M, Rezaee R, Amyot D, Moutthan A, Chreyh R, Geiger G, et al. Real-time and location-based hand hygiene monitoring and notification: proof-of-concept system and experimentation. Pers Ubiquit Comput 2015 Jun 16;19(3-4):667-688. [doi: 10.1007/s00779-015-0855-y]
73. Boudjema S, Dufour JC, Aladro AS, Desquerres I, Brouqui P. MediHandTrace: a tool for measuring and understanding hand hygiene adherence. Clin Microbiol Infect 2014 Jan;20(1):22-28 [FREE Full text] [doi: 10.1111/1469-0691.12471] [Medline: 24261513]
74. Boyce JM, Cooper T, Yin J, Li F, Arbogast JW. Challenges encountered and lessons learned during a trial of an electronic hand hygiene monitoring system. Am J Infect Control 2019 Dec;47(12):1443-1448. [doi: 10.1016/j.ajic.2019.05.019] [Medline: 31324492]
75. Brouqui P, Boudjema S, Soto Aladro A, Chabrière E, Florea O, Nguyen H, et al. New approaches to prevent healthcare-associated infection. Clin Infect Dis 2017 Aug 15;65(suppl_1):50-54. [doi: 10.1093/cid/cix433] [Medline: 28859352]
76. Decker AS, Cipriano GC, Tsouri G, Lavigne JE. Monitoring pharmacy student adherence to World Health Organization hand hygiene indications using radio frequency identification. Am J Pharm Educ 2016 Apr 25;80(3):51 [FREE Full text] [doi: 10.5688/ajpe80351] [Medline: 27170822]
77. Dyson J, Madeo M. Investigating the use of an electronic hand hygiene monitoring and prompt device: influence and acceptability. J Infect Prev 2017 Nov 04;18(6):278-287 [FREE Full text] [doi: 10.1177/1757177417714045] [Medline: 29344097]
78. Filho MA, Marra AR, Magnus TP, Rodrigues RD, Prado M, de Souza Santini TR, et al. Comparison of human and electronic observation for the measurement of compliance with hand hygiene. Am J Infect Control 2014 Nov;42(11):1188-1192. [doi: 10.1016/j.ajic.2014.07.031] [Medline: 25234045]
79. Fisher DA, Seetoh T, May-Lin HO, Viswanathan S, Toh Y, Yin WC, et al. Automated measures of hand hygiene compliance among healthcare workers using ultrasound: validation and a randomized controlled trial. Infect Control Hosp Epidemiol 2013 Sep 02;34(9):919-928. [doi: 10.1086/671738] [Medline: 23917905]
80. Fries J, Segre AM, Thomas G, Herman T, Ellingson K, Polgreen PM. Monitoring hand hygiene via human observers: how should we be sampling? Infect Control Hosp Epidemiol 2012 Jul 02;33(7):689-695 [FREE Full text] [doi: 10.1086/666346] [Medline: 22669230]
81. Herman T, Pemmaraju S, Segre AM, Polgreen PM, Curtis DE, Fries J, et al. Wireless applications for hospital epidemiology. In: Proceedings of the 1st ACM international workshop on Medical-grade wireless networks. 2009 Presented at: MobiHoc '09: The Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing; May 18, 2009; New Orleans Louisiana USA p. 45-50. [doi: 10.1145/1540373.1540384]
82. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. J Infect Dis 2012 Nov 15;206(10):1549-1557 [FREE Full text] [doi: 10.1093/infdis/jis542] [Medline: 23045621]
83. Jain S, Mane S, Lopez J, Lie D, Dallas T, Dissanaike S, et al. A low-cost custom HF RFID system for hand washing compliance monitoring. In: Proceedings of the IEEE 8th International Conference on ASIC. 2009 Presented at: IEEE 8th International Conference on ASIC; Oct. 20-23, 2009; Changsha, China. [doi: 10.1109/asicon.2009.5351530]
84. Johnson R, Tsouri GR, Walsh E. Continuous and automated measuring of compliance of hand-hygiene procedures using finite state machines and RFID. IEEE Instrum Meas Mag 2012 Apr;15(2):8-12. [doi: 10.1109/mim.2012.6174572]
85. Karimpour N, Karaduman B, Ural A, Challengerl M, Dagdeviren O. IoT based hand hygiene compliance monitoring. In: Proceedings of the International Symposium on Networks, Computers and Communications (ISNCC). 2019 Presented at: International Symposium on Networks, Computers and Communications (ISNCC); June 18-20, 2019; Istanbul, Turkey. [doi: 10.1109/isncc.2019.8909151]
86. Levchenko AI, Boscart VM, Ibbett JP, Fernie GR. Distributed IR based technology to monitor hand hygiene of healthcare staff. In: Proceedings of the IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH). 2009 Presented at: IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH); Sept. 26-27, 2009; Toronto, ON, Canada. [doi: 10.1109/tic-sth.2009.5444496]
87. Levchenko AI, Boscart VM, Fernie GR. Automated monitoring: a potential solution for achieving sustainable improvement in hand hygiene practices. Comput Inform Nurs 2014 Aug;32(8):397-403. [doi: 10.1097/CIN.0000000000000067] [Medline: 24859431]

88. Levchenko AI, Boscart VM, Fernie GR. Hand hygiene monitoring and real-time prompting system. In: Proceedings of the IEEE International Systems Conference SysCon 2012. 2012 Presented at: IEEE International Systems Conference SysCon 2012; March 19-22, 2012; Vancouver, BC, Canada. [doi: 10.1109/syscon.2012.6189435]
89. Levchenko AI, Boscart VM, Fernie GR. The effect of automated monitoring and real-time prompting on nurses' hand hygiene performance. Comput Inform Nurs 2013 Oct;31(10):498-504. [doi: 10.1097/01.NCN.0000432124.53902.1d] [Medline: 23924823]
90. Levin PD, Razon R, Schwartz C, Avidan A, Sprung CL, Moses AE, et al. Obstacles to the successful introduction of an electronic hand hygiene monitoring system, a cohort observational study. Antimicrob Resist Infect Control 2019 Feb 22;8(1):43 [FREE Full text] [doi: 10.1186/s13756-019-0498-2] [Medline: 30834111]
91. Marques R, Gregório J, Pinheiro F, Póvoa P, da Silva MM, Lapão LV. How can information systems provide support to nurses' hand hygiene performance? Using gamification and indoor location to improve hand hygiene awareness and reduce hospital infections. BMC Med Inform Decis Mak 2017 Jan 31;17(1):15 [FREE Full text] [doi: 10.1186/s12911-017-0410-z] [Medline: 28143613]
92. Polgreen PM, Hlady CS, Severson MA, Segre AM, Herman T. Method for automated monitoring of hand hygiene adherence without radio-frequency identification. Infect Control Hosp Epidemiol 2010 Dec 02;31(12):1294-1297 [FREE Full text] [doi: 10.1086/657333] [Medline: 20973724]
93. Meydanci MA, Adali C, Ertas M, Dizbay M, Akan A. RFID based hand hygiene compliance monitoring station. In: IEEE International Conference on Control System, Computing and Engineering. 2013 Presented at: IEEE International Conference on Control System, Computing and Engineering; Nov. 29 - Dec. 1, 2013; Penang, Malaysia. [doi: 10.1109/iccsce.2013.6720030]
94. Misra P, Rajaraman V, Aishwarya SN, Dwivedi B, Warrior J. CleanHands: an integrated monitoring system for control of hospital acquired infections. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks. 2015 Presented at: IPSN '15: The 14th International Conference on Information Processing in Sensor Networks; April 13 - 16, 2015; Seattle Washington p. 348-349. [doi: 10.1145/2737095.2742928]
95. Monsalve M, Pemmaraju S, Polgreen PM. Interactions in an intensive care unit: experiences pre-processing sensor network data. In: Proceedings of the 4th Conference on Wireless Health. 2013 Presented at: WH '13: Wireless Health 2013; November 1 - 3, 2013; Baltimore Maryland p. 1-8. [doi: 10.1145/2534088.2534105]
96. Monsalve MN, Pemmaraju SV, Thomas GW, Herman T, Segre AM, Polgreen PM. Do peer effects improve hand hygiene adherence among healthcare workers? Infect Control Hosp Epidemiol 2014 Oct 10;35(10):1277-1285 [FREE Full text] [doi: 10.1086/678068] [Medline: 25203182]
97. Pleteršek A, Sok M, Trontelj J. Monitoring, control and diagnostics using RFID infrastructure. J Med Syst 2012 Dec 22;36(6):3733-3739. [doi: 10.1007/s10916-012-9846-y] [Medline: 22438102]
98. Pong S, Holliday P, Fernie G. Effect of intermittent deployment of an electronic monitoring system on hand hygiene behaviors in healthcare workers. Am J Infect Control 2019 Apr;47(4):376-380 [FREE Full text] [doi: 10.1016/j.ajic.2018.08.029] [Medline: 30502113]
99. Radhakrishna K, Waghmare A, Ekstrand M, Raj T, Selvam S, Sreerama SM, et al. Real-time feedback for improving compliance to hand sanitization among healthcare workers in an open layout ICU using radiofrequency identification. J Med Syst 2015 Jun 10;39(6):68 [FREE Full text] [doi: 10.1007/s10916-015-0251-1] [Medline: 25957165]
100. Srigley JA, Furness CD, Baker GR, Gardam M. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. BMJ Qual Saf 2014 Dec 07;23(12):974-980 [FREE Full text] [doi: 10.1136/bmjqs-2014-003080] [Medline: 25002555]
101. Wan Y, Stombaugh TS, Wang N. Human-subject tracking and localization for a hand hygiene monitoring system. In: Proceedings of the Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS). 2014 Presented at: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS); Nov. 20-21, 2014; Corpus Christi, TX, USA. [doi: 10.1109/upinlbs.2014.7033719]
102. Camilus S, Lee YM. Automated hand hygiene compliance monitoring. In: Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing. 2020 Presented at: IVSP '20: 2020 2nd International Conference on Image, Video and Signal Processing; March 20 - 22, 2020; Singapore p. 64-69. [doi: 10.1145/3388818.3388824]
103. Lacey G, Zhou J, Li X, Craven C, Gush C. The impact of automatic video auditing with real-time feedback on the quality and quantity of handwash events in a hospital setting. Am J Infect Control 2020 Feb;48(2):162-166 [FREE Full text] [doi: 10.1016/j.ajic.2019.06.015] [Medline: 31358419]
104. Zhong H, Kanhere SS, Chou CT. WashInDepth: Lightweight hand wash monitor using depth sensor. In: Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. 2016 Presented at: MOBIQUITOUS 2016: Computing, Networking and Services; November 28 - December 1, 2016; Hiroshima Japan p. 28-37. [doi: 10.1145/2994374.2994386]
105. Khamis A, Kusy B, Chou CT, McLaws M, Hu W. Poster abstract: a weakly supervised tracking of hand hygiene technique. In: Proceedings of the 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). 2020 Presented at: 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN); April 21-24, 2020; Sydney, NSW, Australia. [doi: 10.1109/ipsn48710.2020.00-21]

106. Khan A, Nausheen S. Compliance of surgical hand washing before surgery: role of remote video surveillance. J Pak Med Assoc 2017 Jan;67(1):92-96 [FREE Full text] [Medline: 28065962]
107. Banerjee A, Amperyani VN, Gupta SK. Hand hygiene compliance checking system with explainable feedback. In: Proceedings of the 6th ACM Workshop on Wearable Systems and Applications. 2020 Presented at: MobiSys '20: The 18th Annual International Conference on Mobile Systems, Applications, and Services; June 19, 2020; Toronto Ontario Canada p. 34-36. [doi: 10.1145/3396870.3400015]
108. Kutafina E, Laukamp D, Jonas SM. Wearable sensors in medical education: supporting hand hygiene training with a forearm EMG. Stud Health Technol Inform 2015;211:286-291. [Medline: 25980884]
109. Kutafina E, Laukamp D, Bettermann R, Schroeder U, Jonas SM. Wearable sensors for elearning of manual tasks: using forearm EMG in hand hygiene training. Sensors (Basel) 2016 Aug 03;16(8):1221 [FREE Full text] [doi: 10.3390/s16081221] [Medline: 27527167]
110. Wang C, Sarsenbayeva Z, Chen X, Dingler T, Goncalves J, Kostakos V. Accurate measurement of handwash quality using sensor armbands: instrument validation study. JMIR Mhealth Uhealth 2020 Mar 26;8(3):e17001 [FREE Full text] [doi: 10.2196/17001] [Medline: 32213469]
111. Pellegrino R, Crandall PG, Seo HS. Using olfaction and unpleasant reminders to reduce the intention-behavior gap in hand washing. Sci Rep 2016 Jan 06;6(1):18890 [FREE Full text] [doi: 10.1038/srep18890] [Medline: 26732033]
112. Srisomboon K, Malathum K, Skuntaniyom S, Duangthongkham J, Phanthuna N. Hand hygiene auditing algorithm using image processing for hand hygiene evaluation machine. In: Proceedings of the 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). 2019 Presented at: 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON); July 10-13, 2019; Pattaya, Thailand. [doi: 10.1109/ecti-con47248.2019.8955358]
113. Szilágyi L, Lehotsky A, Nagy M, Haidegger T, Benyó B, Benyó Z. Stery-hand: a new device to support hand disinfection. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; Aug. 31 - Sept. 4, 2010; Buenos Aires, Argentina. [doi: 10.1109/iembs.2010.5626377]
114. Yamamoto K, Miyanaga K, Miyahara H, Yoshii M, Kinoshita F, Touyama H. Toward the evaluation of handwashing skills based on image processing. In: Proceedings of the Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS). 2018 Presented at: Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS); Dec. 5-8, 2018; Toyama, Japan. [doi: 10.1109/scis-isis.2018.00141]
115. Yamamoto K, Yoshii M, Kinoshita F, Touyama H. Classification vs regression by CNN for handwashing skills evaluations in nursing education. In: International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2020 Presented at: International Conference on Artificial Intelligence in Information and Communication (ICAIIC); Feb. 19-21, 2020; Fukuoka, Japan. [doi: 10.1109/icaiic48513.2020.9064974]
116. Naim F, Jaafar R, Arshad NW, Hamid R, Razali M. Unclean hand detection machine using vision sensor. In: Proceedings of the Saudi International Electronics, Communications and Photonics Conference. 2013 Presented at: Saudi International Electronics, Communications and Photonics Conference; April 27-30, 2013; Riyadh, Saudi Arabia. [doi: 10.1109/siecpc.2013.6550750]
117. Hlady CS, Severson MA, Segre AM, Polgreen PM. A mobile handheld computing application for recording hand hygiene observations. Infect Control Hosp Epidemiol 2010 Sep 02;31(9):975-977 [FREE Full text] [doi: 10.1086/655834] [Medline: 20636134]
118. Kinney P. ZigBee Technology: wireless control that simply works. Commun Des Conf. 2003. URL: https://www.itk.ntnu.no/fag/TTK4545/TTK2/PDF/ZigBee%20Technology%20-%20Wireless%20control%20that%20simply%20works.pdf [accessed 2021-10-22]
119. Goroncy-Bermes P, Koburger T, Meyer B. Impact of the amount of hand rub applied in hygienic hand disinfection on the reduction of microbial counts on hands. J Hosp Infect 2010 Mar;74(3):212-218. [doi: 10.1016/j.jhin.2009.09.018] [Medline: 20061058]
120. Lehotsky A, Szilágyi L, Bánsághi S, Szerémy P, Wéber G, Haidegger T. Towards objective hand hygiene technique assessment: validation of the ultraviolet-dye-based hand-rubbing quality assessment procedure. J Hosp Infect 2017 Sep;97(1):26-29. [doi: 10.1016/j.jhin.2017.05.022] [Medline: 28579470]
121. Conway LJ. Challenges in implementing electronic hand hygiene monitoring systems. Am J Infect Control 2016 May 02;44(5 Suppl):7-12. [doi: 10.1016/j.ajic.2015.11.031] [Medline: 27131139]
122. Ellingson K, Polgreen PM, Schneider A, Shinkunas L, Kaldjian LC, Wright D, et al. Healthcare personnel perceptions of hand hygiene monitoring technology. Infect Control Hosp Epidemiol 2011 Nov 02;32(11):1091-1096. [doi: 10.1086/662179] [Medline: 22011536]
123. Gu Y, Lo A, Niemegeers I. A survey of indoor positioning systems for wireless personal networks. IEEE Commun Surv Tutorials 2009;11(1):13-32. [doi: 10.1109/surv.2009.090103]

124. Fung ML, Chen MZ, Chen YH. Sensor fusion: A review of methods and applications. In: Proceedings of the 29th Chinese Control And Decision Conference (CCDC). 2017 Presented at: 29th Chinese Control And Decision Conference (CCDC); May 28-30, 2017; Chongqing, China. [doi: 10.1109/ccdc.2017.7979175]
125. Wang C, Sarsenbayeva Z, Luo C, Goncalves J, Kostakos V. Improving wearable sensor data quality using context markers. In: Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. 2019 Presented at: UbiComp '19: The 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 9 - 13, 2019; London United Kingdom p. 598-601. [doi: 10.1145/3341162.3349334]
126. Bannach D, Amft O, Lukowicz P. Automatic event-based synchronization of multimodal data streams from wearable and ambient sensors. In: Barnaghi P, Moessner K, Presser M, Meissner S, editors. Smart Sensing and Context. Berlin, Heidelberg: Springer; 2009:135-148.
127. Palmore T, Henderson DK. Big brother is washing...Video surveillance for hand hygiene adherence, through the lenses of efficacy and privacy. Clin Infect Dis 2012 Jan 01;54(1):8-9 [FREE Full text] [doi: 10.1093/cid/cir781] [Medline: 22109949]
128. Cilliers L. Wearable devices in healthcare: privacy and information security issues. Health Inf Manag 2020 May 30;49(2-3):150-156. [doi: 10.1177/1833358319851684] [Medline: 31146589]
129. American Conference of Governmental Industrial Hygienists. TLVs and BEIs : Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices. Cincinnati, OH: ACGIH Worldwide; 2016.
130. Zarb P, Coignard B, Griskeviciene J, Muller A, Vankerckhoven V, Weist K, National Contact Points for the ECDC pilot point prevalence survey, Hospital Contact Points for the ECDC pilot point prevalence survey. The European Centre for Disease Prevention and Control (ECDC) pilot point prevalence survey of healthcare-associated infections and antimicrobial use. Euro Surveill 2012 Nov 15;17(46):20316 [FREE Full text] [doi: 10.2807/ese.17.46.20316-en] [Medline: 23171822]
131. Trick WE, Vernon MO, Hayes RA, Nathan C, Rice TW, Peterson BJ, et al. Impact of ring wearing on hand contamination and comparison of hand hygiene agents in a hospital. Clin Infect Dis 2003 Jun 01;36(11):1383-1390. [doi: 10.1086/374852] [Medline: 12766832]
132. Bassen HI. Radiofrequency interference with medical devices. A technical information statement. IEEE Eng Med Biol Mag 1998;17(3):111-114. [Medline: 9604711]
133. Badizadegan ND, Greenberg S, Lawrence H, Badizadegan K. Radiofrequency interference in the clinical laboratory. Am J Clin Pathol 2019 Apr 02;151(5):522-528 [FREE Full text] [doi: 10.1093/ajcp/aqy174] [Medline: 30668626]
134. van der Togt R, van Lieshout EJ, Hensbroek R, Beinat E, Binnekade JM, Bakker PJ. Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment. J Am Med Assoc 2008 Jun 25;299(24):2884-2890. [doi: 10.1001/jama.299.24.2884] [Medline: 18577733]
135. Li X, Zhang D, Lv Q, Xiong J, Li S, Zhang Y, et al. IndoTrack: Device-free indoor human tracking with commodity Wi-Fi. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-22. [doi: 10.1145/3130940]
136. Matteson DS, James NA. A nonparametric approach for multiple change point analysis of multivariate data. J Am Stat Assoc 2014 Mar 19;109(505):334-345. [doi: 10.1080/01621459.2013.849605]
137. Fasching J, Walczak N, Bernstein GA, Hadjiyanni T, Cullen K, Morellas V, et al. Automated coding of activity videos from an OCD study. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). 2016 Presented at: IEEE International Conference on Robotics and Automation (ICRA); May 16-21, 2016; Stockholm, Sweden. [doi: 10.1109/icra.2016.7487783]
138. Ashraf A, Taati B. Automated video analysis of handwashing behavior as a potential marker of cognitive health in older adults. IEEE J Biomed Health Inform 2016 Mar;20(2):682-690. [doi: 10.1109/jbhi.2015.2413358]
139. Barnes S, Golden B, Wasil E, Furuno J, Harris A. An application of factorial design to compare the relative effectiveness of hospital infection control measures. In: Proceedings of the 2011 Winter Simulation Conference (WSC). 2011 Presented at: Winter Simulation Conference (WSC); Dec. 11-14, 2011; Phoenix, AZ, USA. [doi: 10.1109/wsc.2011.6147849]

Abbreviations

BLE: Bluetooth low energy
HAI: health care–associated infections
HCW: health care worker
IMU: inertial measurement unit
LED: light-emitting diode
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RFI: radio-frequency interference
RFID: radio-frequency identification
RGB: red green blue
RTLS: real-time locating system
sEMG: surface electromyography
WHO: World Health Organization

Edited by R Kukafka, G Eysenbach; submitted 11.02.21; peer-reviewed by Y Liu, H Mehdizadeh, S Gaube; comments to author 17.08.21; revised version received 04.09.21; accepted 04.10.21; published 24.11.21.

Please cite as:
Wang C, Jiang W, Yang K, Yu D, Newn J, Sarsenbayeva Z, Goncalves J, Kostakos V
Electronic Monitoring Systems for Hand Hygiene: Systematic Review of Technology
J Med Internet Res 2021;23(11):e27880
URL: https://www.jmir.org/2021/11/e27880
doi:10.2196/27880
PMID:34821565

Viewpoint

Sensor Data Integration: A New Cross-Industry Collaboration to Articulate Value, Define Needs, and Advance a Framework for Best Practices

Ieuan Clay[1], PhD; Christian Angelopoulos[2], BSc; Anne Lord Bailey[3], PharmD, BCPS; Aaron Blocker[4], MSc; Simona Carini[5], MA; Rodrigo Carvajal[6], BSc; David Drummond[7], PhD; Kimberly F McManus[3], MSc, PhD; Ingrid Oakley-Girvan[8], MPH, PhD; Krupal B Patel[6], MD, MSc; Phillip Szepietowski[6], PhD; Jennifer C Goldsack[1], MChem, MA, MBA

[1]Digital Medicine Society (DiMe), Boston, MA, United States
[2]HumanFirst, San Fransisco, CA, United States
[3]Department of Veterans Affairs, Washington, DC, United States
[4]Savvy Cooperative, Queens, NY, United States
[5]University of California San Francisco, San Francisco, CA, United States
[6]H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, United States
[7]Evidation Health, San Mateo, CA, United States
[8]Medable, San Fransisco, CA, United States

Corresponding Author:
Ieuan Clay, PhD
Digital Medicine Society (DiMe)
90 Canal Street
Boston, MA, 02114
United States
Phone: 1 765 234 3463
Email: ieuan@dimesociety.org

Abstract

Data integration, the processes by which data are aggregated, combined, and made available for use, has been key to the development and growth of many technological solutions. In health care, we are experiencing a revolution in the use of sensors to collect data on patient behaviors and experiences. Yet, the potential of this data to transform health outcomes is being held back. Deficits in standards, lexicons, data rights, permissioning, and security have been well documented, less so the cultural adoption of sensor data integration as a priority for large-scale deployment and impact on patient lives. The use and reuse of trustworthy data to make better and faster decisions across drug development and care delivery will require an understanding of all stakeholder needs and best practices to ensure these needs are met. The Digital Medicine Society is launching a new multistakeholder Sensor Data Integration Tour of Duty to address these challenges and more, providing a clear direction on how sensor data can fulfill its potential to enhance patient lives.

(J Med Internet Res 2021;23(11):e34493) doi:10.2196/34493

KEYWORDS

digital measures; data integration; patient centricity; utility

Introduction

Data integration has been defined as "...the technical and business processes used to (aggregate and) combine data from multiple sources to provide a unified, single view of the data" [1]. In practice, this means building bridges across different sources of data, to create an integrated data set that enables you to ask bigger questions or new questions (ie, integration followed by utilization) [2].

Data integration has been a force for innovation and scaling in many technological fields, helping new products become part of our everyday lives. Why do your new headphones work with your phone? Cross-industry collaboration on standards fueled the growth of Bluetooth from its initial invention at Ericsson in the late 1990s to one of the most prevalent current-day

technologies [3], incorporated into an estimated 5 billion new devices in 2021 [4]. Love that your credit card works practically everywhere, including internationally? This would not be possible without the Payment Card Industry Data Security Standard (and associated Security Council), which has been a globally accepted standard since 2004 [5]. Trying to get home the fastest way? The apps and services that help us plan and navigate our journeys rely on the General Transit Feed Specification [6], which began in 2005 as a collaboration between Google and the TriMet transport agency and is now a global standard driving the utilization of transit data [7]. All of these examples underline not only the practical impact of successful data integration, but also the critical role of multisector collaboration in achieving success.

Where is the new field of digital health, and sensor data in particular, in its data integration journey and what needs to be done to move it forward?

The Digital Medicine Society (DiMe), a nonprofit organization dedicated to advancing the safe, effective, ethical, and equitable use of digital medicine to optimize health [8], will examine these questions and more as part of a new Sensor Data Integration Tour of Duty. The tour will convene multistakeholder experts from Amazon Web Services, Evidation Health, the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) and the Center for Devices and Radiological Health (CDRH), Human First, the Institute of Electrical and Electronics Engineers, Medable, the Moffitt Cancer Center, Oracle, Open mHealth, Savvy Cooperative, Takeda, and the United States Department of Veterans Affairs. The Sensor Data Integration Tour of Duty will build off other initiatives like The Playbook [9] in advancing the adoption and utility of digital medicine tools. In this paper, we will describe the current state of sensor data integration in health care, examine the importance of sensor data integration in achieving the promise of digital health to improve clinical research and patient care at scale, and offer a vision of what overcoming these challenges and continuing progress will enable.

Why Is Sensor Data Integration Important?

In the world of health care, examples of data integration can be found across the drug development and care delivery continuum:

- Modeling and understanding complex disease systems is typically only possible by considering and integrating a wide range of relevant data sources [10,11];
- Data integration supports translating observations from preclinical models to studies with human subjects and is common in early phase drug discovery [12,13];
- Similar data sources combined across different studies create meta-studies to increase statistical power to detect differences and improve the generalizability of results [14,15];
- Bridging clinical to real-world evidence studies aids the translation of observations from clinical development to the market. Examples include converting an outcome used in a clinical trial to deployment to support reimbursement [16] or treatment [17], or comparing error from research or clinical-grade and consumer-grade sensors to better validate large-scale observations [18].

Data integration is ubiquitous in digital health, but the definition above highlights three critical considerations pertaining to sensor data from digital health products:

- Solutions to sensor data integration challenges will be found in both “technical and business processes”;
- “Aggregating and combining data from multiple sources” will provide the critical context for sensor data to drive a learning health care system [19];
- Creating a “unified, single view of the data” will allow the use and reuse of trustworthy data to make better and faster decisions.

Sensor data (ie, data collected by wearable devices, smartphones, and other sensor-equipped connected digital medicine products [20]) is not a silver bullet. We cannot develop new sensor-based measures without combining data from different data sources [21,22], and we cannot deploy those measures without contextualization [23,24].

When we develop a sensor-based digital measure, establishing that it is fit for purpose in a given population involves cross-referencing with clinical and behavioral anchor data to demonstrate accuracy and validity [20,25]. As a therapy moves from clinical development into the hands of patients, it is also critical that clinical outcomes are integrated with real-world data, and that researcher stakeholders and other decision-makers such as payers and clinical policy makers can access and assess those data.

To support utilization and full patient value, health care providers and hospital administrators making treatment decisions all need access to the sensor data and digital measures captured during care and need to be able to digest and infer the same conclusions. Similarly, participants, investigators, statisticians, heads of medical affairs, or regulators need those data to be contextualized and appropriately presented to them, such that they are able to draw insights and value.

Currently, we are far from achieving this potential to improve lives, and it has been claimed that limitations in data integration and utilization are the primary rate limiter [26,27]. Improving lives is always at the front of our minds, making it worthwhile to think about how these tools and applications “look” from the patient’s perspective. Data integration should minimize additional burden caused by management and assessment tools to patients already burdened by their condition(s). Widespread adoption will require interfaces that provide immediate value and are simple, interactive, and intuitive. No longer should patients ask: why does my health care provider not have access to all relevant information? Why should I have to perform the same test in multiple trials, and why do they all have separate apps? I think this data is important; how do I share it with my doctor? Achieving truly integrated continuous care is not possible without effective data integration [28].

What Are the Key Issues?

There are several specific technical hurdles that get in the way of effective data integration, including the plethora of formats [29,30] and the corresponding lack of consensus around common metadata standards [14,26], although progress is being made [31]. Perhaps more worrisome is the fact that permissioning and consent often exclude integration and exploration [32], and massive gaps in digital health research remain regarding data rights and governance [33,34].

Yet, perhaps the biggest hurdles are cultural adoption, that is, effective incentivization to share data (within and across organizations) and make it easier to integrate [35-37]; recognition that data integration, while perhaps not a glorious pursuit, is central to scaling any digital health tool [38]; and lack of a unified signal from clinical (and other) stakeholders that would enable vendors to address needs around data integration [39,40].

Data integration for digital health, and for sensor data specifically, is in the early phases, and we have yet to see the kind of progress that has been made in other data-heavy health fields (eg, genomics) [41-44]. Currently, no widely adopted data standards or repositories exist, and while progress is being made on the integration of other data types (eg, electronic health records [EHRs], molecular data, patient reported outcomes), we see limited progress around sensor data specifically. This presents an opportunity to provide an early framework for sensor-based data integration and future success of this technology and avoid barriers in integrating other health care data (eg, EHRs). Moreover, understanding the legal and ethical consequences of sharing and integrating sensor data is at a nascent stage [45]. Recent work has shown that "the three most underrepresented areas of research into digital clinical measures were ethics, security, and data rights and governance" [33].

Where Has Progress Been Made?

Data integration has been recognized as a key driver of value creation in the private sector. In the past few years, we have seen companies invest to improve their data integration abilities through smaller venture capital deals [46], larger acquisitions [47], or internal initiatives like data42 from Novartis [48].

The public space has also been very active, perhaps helped by the inherently precompetitive nature of such work (ie, progress in data integration helps all parties, without any individual party needing to share proprietary data). Across research, clinical development, and in real-world practice, we are starting to see encouraging progress.

Research has been boosted by a number of initiatives, often focused on improving the awareness and availability of digital health data sets. Examples include meta-reviews of publicly available data sets in specific fields to accelerate integration and use of these data, as well as highlighting potential issues and shortcomings like demographic and ethnic underrepresentation [14]. Similarly, new repositories and portals are helping researchers connect and make their data available, such as Synapse and their Digital Health portal [49] or Zenodo [50]. Zenodo is a robust example as it grew out of OpenAIRE [51], a public project focused on enabling open science, demonstrating the concrete progress that specific projects can drive. The COVID-19 Evidence Accelerator [52] has also contributed to integration across the colossal volumes of data being developed as part of the pandemic response.

Clinical development and care delivery are also benefiting from public initiatives around data standards like Substitutable Medical Applications, Reusable Technologies on Fast Healthcare Interoperability Resources (SMART on FHIR) [53], which focuses on EHRs, or the Clinical Data Interchange Standards Consortium, which has had a substantial impact on the submission of evidence to regulatory bodies [54]. RADAR-base (Remote Assessment of Disease and Relapse) [55] grew out of the Innovative Medicine Initiative's RADAR-CNS (RADAR in Central Nervous System Disorders) [56] and aims to generalize data integration tools and best practices. These initiatives focus both on the integration of a range of devices and data sources, including mobile apps, and the utilization of the data through standardized visualization and analysis tools. Additional examples are the data integration centers in German medical universities, which have been set up to enable use of data across research and care [57]. Establishing these centers is a "meta-collaboration" of 4 consortia funded by the German Ministry of Education and Research. The recently formed Graphite Health is a nonprofit creating solutions for better health system interoperability [58].

Subsequent data utilization efforts bring data rights to the forefront. Who should be able to use the integrated data sets, and for what purpose? How can consenting and integration be structured such that innovation is not hindered and we remain within the ethical boundaries protecting patient rights? The FDA Sentinel Initiative combines claims data from multiple insurers, data from EHRs and patient reports provided by multiple health systems, and even some sensor technologies to evaluate the safety of medical products in the real world [59]. It combines a common data model with a distributed data architecture. Approved users are given secure, structured access to digital health data while respecting privacy and proprietary concerns. The Sentinel Initiative thus protects the data rights of organizations and individuals by keeping data with the holder while still allowing that data to be used for beneficial (eg, research) purposes. Similarly, the SELFIE Horizon 2020 consortium [60] also puts data rights at the forefront, recognizing that data integration is a key pillar of patient-centric, integrated, continuous care, while attempting to balance "preservation of individual privacy and the need for health data sharing" [28].

What does all this mean for sensor data? To achieve real impact on patient lives, sensor data needs to be integrated and contextualized with many other data types. Concerted progress must therefore be made across digital health, and sensor data must keep up with other sources of data like EHRs and break out of condition-specific efforts.

A Vision for the Future

Sensor data that is accessible, trustworthy, and relevant

Overcoming data integration challenges is immeasurably worthwhile. If we are to realize the potential of these new data flows to improve patient lives, then we will move toward integrated, holistic, and continuous care [28]. We will provide a smoother, less burdensome, and more valuable experience for patients. We will accelerate research and innovation, and enable the incorporation of data collected from everyday experiences to be included in decision-making. We will dovetail with more mature efforts across health care and do this using a collaborative, ethical, and multisector approach that will ensure that the potential of this data is realized for all stakeholders.

Imagine a situation where a diabetes patient is diagnosed with chronic obstructive pulmonary disease (COPD). The patient can log medication use via an app, and securely and privately funnel activity and sleep data from their user-friendly smartwatch into their doctor's health care systems. This data is of sufficient quality and accompanied by enough contextual information to facilitate the conversation between patient and doctor ("how have you been feeling?" becomes "I see you have been sleeping better; do you also feel better?"), and enable better shared decision-making. Exacerbations are tracked and alerts are shared with the patient's caregivers, which gives the patient peace of mind, and positive reinforcement to further improve this tool. The patient consents that researchers can access their data to improve the detection algorithm, and this data is later reused by a cross-industry consortium to drive qualification of a new COPD outcome.

In oncology settings, the potential is equally appealing. Passive biometric data provides additional data points on a patient's physical well-being, complementing patient reported outcomes and other data. Worrisome symptoms and activities can be triaged and brought to the clinical care team's attention. With just-in-time interventions, treatment toxicities can be mitigated or managed early, reducing the downstream costs of readmissions or emergency room visits. Simple interfaces that help patients track their data and symptoms would provide positive reinforcement to support behavior changes. Properly integrated, this data can also help industry partners such as pharmaceutical companies track outcomes and treatment toxicities in clinical trials.

A future where real-world data helps drive the health care system ultimately for preventive actions versus reactive care is a bright one. It is an environment where clinical research and care are not siloed and the data we generate can be referenced again and again, with appropriate consent, to reimagine a system informed by our physiology and how we feel, function, and survive.

Outlook

Integration and utilization of sensor data as part of the wider landscape of health-relevant data is still in its infancy, but progress is being made. Resolving current shortcomings will reduce burden to, and better serve, all stakeholders: patients, caregivers, clinicians, health care administrators, payers, industry partners, technology manufacturers, analytics companies, and researchers. The Sensor Data Integration Tour of Duty from DiMe includes experts from Amazon Web Services, Evidation Health, the FDA CDER and CDRH, Human First, the Institute of Electrical and Electronics Engineers, Medable, the Moffitt Cancer Center, Oracle, Open mHealth, Savvy Cooperative, Takeda, and the United States Department of Veterans Affairs. Our focus will be to examine stakeholder needs and define a framework for integrating sensor data and making it more useful for decision-making within the wider health care landscape.

Sign up as a DiMe member to stay up to date on findings and resources from this and other projects [8], and check the DiMe e-collection of articles for relevant papers at https://jmir.org/themes/1160-digital-medicine-society-dime.

Acknowledgments

The authors would like to thank all the Sensor Data Integration Tour of Duty participants for their inspiring ideas, energy, and enthusiasm.

Conflicts of Interest

None declared.

References

1. Data Integration. OmniSci, Inc. 2021. URL: https://www.omnisci.com/technical-glossary/data-integration [accessed 2021-09-14]
2. Peng C, Goswami P, Bai G. A literature review of current technologies on health data integration for patient-centered health management. Health Informatics J 2020 Sep;26(3):1926-1951 [FREE Full text] [doi: 10.1177/1460458219892387] [Medline: 31884843]
3. Haartsen J. How we made Bluetooth. Nat Electron 2018 Dec 13;1(12):661-661. [doi: 10.1038/s41928-018-0186-x]
4. Zeadally S, Siddiqui F, Baig Z. 25 Years of Bluetooth Technology. Future Internet 2019 Sep 09;11(9):194. [doi: 10.3390/fi11090194]
5. PCI Security. PCI Security Standards Council. URL: https://www.pcisecuritystandards.org/pci_security/ [accessed 2021-09-16]

6. GTFS: Making Public Transit Data Universally Accessible. General Transit Feed Specification. URL: https://gtfs.org/ [accessed 2021-09-16]
7. Chapter 10 Pioneering Open Data Standards: The GTFS Story. Beyond Transparency. URL: https://beyondtransparency.org/chapters/part-2/pioneering-open-data-standards-the-gtfs-story/ [accessed 2021-09-16]
8. Digital Medicine Society (DiMe). Advancing digital medicine to optimize human health. 2018. URL: https://www.dimesociety.org/ [accessed 2021-09-29]
9. The Playbook. Digital Medicine Society (DiMe). URL: https://playbook.dimesociety.org/ [accessed 2021-10-28]
10. Stephenson D, Alexander R, Aggarwal V, Badawy R, Bain L, Bhatnagar R, et al. Precompetitive Consensus Building to Facilitate the Use of Digital Health Technologies to Support Parkinson Disease Drug Development through Regulatory Science. Digit Biomark 2020;4(Suppl 1):28-49 [FREE Full text] [doi: 10.1159/000512500] [Medline: 33442579]
11. Beattie Z, Miller LM, Almirola C, Au-Yeung WM, Bernard H, Cosgrove KE, et al. The Collaborative Aging Research Using Technology Initiative: An Open, Sharable, Technology-Agnostic Platform for the Research Community. Digit Biomark 2020;4(Suppl 1):100-118 [FREE Full text] [doi: 10.1159/000512208] [Medline: 33442584]
12. Hartl D, de Luca V, Kostikova A, Laramie J, Kennedy S, Ferrero E, et al. Translational precision medicine: an industry perspective. J Transl Med 2021 Jun 05;19(1):245 [FREE Full text] [doi: 10.1186/s12967-021-02910-6] [Medline: 34090480]
13. De Maria Marchiano R, Di Sante G, Piro G, Carbone C, Tortora G, Boldrini L, et al. Translational Research in the Era of Precision Medicine: Where We Are and Where We Will Go. J Pers Med 2021 Mar 18;11(3) [FREE Full text] [doi: 10.3390/jpm11030216] [Medline: 33803592]
14. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. The Lancet Digital Health 2021 Jan;3(1):e51-e66 [FREE Full text] [doi: 10.1016/S2589-7500(20)30240-5] [Medline: 33735069]
15. Mitchell WG, Dee EC, Celi LA. Generalisability through local validation: overcoming barriers due to data disparity in healthcare. BMC Ophthalmol 2021 May 21;21(1):228 [FREE Full text] [doi: 10.1186/s12886-021-01992-6] [Medline: 34020592]
16. Wyner Z, Dublin S, Chambers C, Deval S, Herzig-Marx C, Rao S, et al. The FDA MyStudies app: a reusable platform for distributed clinical trials and real-world evidence studies. JAMIA Open 2020 Dec;3(4):500-505 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa061] [Medline: 33623887]
17. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
18. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
19. Learning Healthcare System. The Learning Healthcare Project. 2017. URL: http://learninghealthcareproject.org/background/learning-healthcare-system/ [accessed 2021-10-15]
20. 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]
21. Mueller A, Hoefling HA, Muaremi A, Praestgaard J, Walsh LC, Bunte O, et al. Continuous Digital Monitoring of Walking Speed in Frail Elderly Patients: Noninterventional Validation Study and Longitudinal Clinical Trial. JMIR Mhealth Uhealth 2019 Nov 27;7(11):e15191 [FREE Full text] [doi: 10.2196/15191] [Medline: 31774406]
22. Servais L, Camino E, Clement A, McDonald C, Lukawy J, Lowes L, et al. First Regulatory Qualification of a Novel Digital Endpoint in Duchenne Muscular Dystrophy: A Multi-Stakeholder Perspective on the Impact for Patients and for Drug Development in Neuromuscular Diseases. Digit Biomark 2021 Aug 5;5(2):183-190. [doi: 10.1159/000517411]
23. Lin YL, Trbovich P, Kolodzey L, Nickel C, Guerguerian A. Association of Data Integration Technologies With Intensive Care Clinician Performance: A Systematic Review and Meta-analysis. JAMA Netw Open 2019 May 03;2(5):e194392 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.4392] [Medline: 31125104]
24. Smart data integration along the patient pathway. Healthcare IT News. 2021 Jul 19. URL: https://www.healthcareitnews.com/news/smart-data-integration-along-patient-pathway [accessed 2021-09-16]
25. 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]
26. Brayanov J, Goldsack J, Byrom B. With Covid-19 halting clinical trials, wearables could be key — but data 'wild west' gets in the way. STAT. 2020 Aug 11. URL: https://www.statnews.com/2020/08/11/with-covid-19-halting-clinical-trials-wearables-could-be-key-but-data-wild-west-gets-in-the-way/ [accessed 2021-09-15]
27. Melchiorre MG, Papa R, Rijken M, van Ginneken E, Hujala A, Barbabella F. eHealth in integrated care programs for people with multimorbidity in Europe: Insights from the ICARE4EU project. Health Policy 2018 Jan;122(1):53-63 [FREE Full text] [doi: 10.1016/j.healthpol.2017.08.006] [Medline: 28899575]

28. Baltaxe E, Czypionka T, Kraus M, Reiss M, Askildsen JE, Grenkovic R, et al. Digital Health Transformation of Integrated Care in Europe: Overarching Analysis of 17 Integrated Care Programs. J Med Internet Res 2019 Sep 26;21(9):e14956 [FREE Full text] [doi: 10.2196/14956] [Medline: 31573914]
29. Ilan Y. Second-Generation Digital Health Platforms: Placing the Patient at the Center and Focusing on Clinical Outcomes. Front Digit Health 2020 Dec 3;2. [doi: 10.3389/fdgth.2020.569178]
30. Eschrich S, Teer J, Reisman P, Siegel E, Challa C, Lewis P, et al. Enabling Precision Medicine in Cancer Care Through a Molecular Data Warehouse: The Moffitt Experience. JCO Clinical Cancer Informatics 2021 Jun(5):561-569. [doi: 10.1200/cci.20.00175]
31. IEEE. 1752.1-2021 - IEEE Standard for Open Mobile Health Data—Representation of Metadata, Sleep, and Physical Activity Measures. IEEE Std 1752.1-2021 2021:1-24 [FREE Full text] [doi: 10.1109/IEEESTD.2021.9540821]
32. Ventresca M, Schünemann HJ, Macbeth F, Clarke M, Thabane L, Griffiths G, et al. Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide. BMC Med Res Methodol 2020 May 12;20(1):113 [FREE Full text] [doi: 10.1186/s12874-020-00964-6] [Medline: 32398016]
33. Shandhi M, Goldsack J, Ryan K, Bennion A, Kotla A, Feng A, et al. Recent Academic Research on Clinically Relevant Digital Measures: Systematic Review. J Med Internet Res 2021 Sep 15;23(9):e29875 [FREE Full text] [doi: 10.2196/29875] [Medline: 34524089]
34. Proudlock D. Quality Generation & Ethical Use of Digital Health Data in Clinical Studies. Pistoia Alliance. 2020. URL: https://www.pistoiaalliance.org/projects/current-projects/digital-health-data-in-clinical-studies/ [accessed 2021-10-11]
35. FAIR Principles. FAIR. 2017. URL: https://www.go-fair.org/fair-principles/ [accessed 2021-09-15]
36. Wilkinson MD, Sansone S, Schultes E, Doorn P, Bonino da Silva Santos LO, Dumontier M. A design framework and exemplar metrics for FAIRness. Sci Data 2018 Jun 26;5:180118 [FREE Full text] [doi: 10.1038/sdata.2018.118] [Medline: 29944145]
37. Koers H, Bangert D, Hermans E, van Horik R, de Jong M, Mokrane M. Recommendations for Services in a FAIR Data Ecosystem. Patterns (N Y) 2020 Aug 14;1(5):100058 [FREE Full text] [doi: 10.1016/j.patter.2020.100058] [Medline: 33205119]
38. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
39. Weber S, Heitmann KU. [Interoperability in healthcare: also prescribed for digital health applications (DiGA)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1262-1268 [FREE Full text] [doi: 10.1007/s00103-021-03414-w] [Medline: 34532746]
40. Abbas A, Schultebraucks K, Galatzer-Levy I. Digital Measurement of Mental Health: Challenges, Promises, and Future Directions. Psychiatric Annals 2021 Jan;51(1):14-20. [doi: 10.3928/00485713-20201207-01]
41. Alterovitz G, Heale B, Jones J, Kreda D, Lin F, Liu L, et al. FHIR Genomics: enabling standardization for precision medicine use cases. NPJ Genom Med 2020;5:13 [FREE Full text] [doi: 10.1038/s41525-020-0115-6] [Medline: 32194985]
42. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. Blood 2017 Jul 27;130(4):453-459 [FREE Full text] [doi: 10.1182/blood-2017-03-735654] [Medline: 28600341]
43. Villanueva AG, Cook-Deegan R, Robinson JO, McGuire AL, Majumder MA. Genomic Data-Sharing Practices. J Law Med Ethics 2019 Mar;47(1):31-40 [FREE Full text] [doi: 10.1177/1073110519840482] [Medline: 30994063]
44. Huang S, Chaudhary K, Garmire LX. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. Front Genet 2017;8:84 [FREE Full text] [doi: 10.3389/fgene.2017.00084] [Medline: 28670325]
45. 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]
46. Press G. Centaur Labs Gets $15 Million To Improve Data For Health care AI. Forbes. 2021 Sep 3. URL: https://www.forbes.com/sites/gilpress/2021/09/03/centaur-labs-gets-15-million-to-improve-data-for-healthcare-ai/?sh=251eb12863a0 [accessed 2021-09-10]
47. Roche completes acquisition of Flatiron Health. F Hoffmann-La Roche Ltd. 2018 Apr 6. URL: https://www.roche.com/media/releases/med-cor-2018-04-06.htm [accessed 2021-09-10]
48. Mijuk G. The data42 program shows Novartis' intent to go big on data and digital. Novartis. 2020 Mar 11. URL: https://www.novartis.com/stories/discovery/data42-program-shows-novartis-intent-go-big-data-and-digital [accessed 2021-09-10]
49. Synapse Digital Health Portal. Synapse. URL: https://dhealth.synapse.org/ [accessed 2021-09-10]
50. Zenodo - Research. URL: https://zenodo.org/ [accessed 2021-09-15]
51. OpenAIRE. URL: https://www.openaire.eu/ [accessed 2021-09-15]
52. COVID-19 Evidence Accelerator. URL: https://www.evidenceaccelerator.org/ [accessed 2021-09-16]
53. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep 17;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
54. CDISC. URL: https://www.cdisc.org/ [accessed 2021-09-29]

55. RADAR-base. 2018. URL: https://radar-base.org/ [accessed 2021-09-29]
56. RADAR-CNS: Remote Assessment of Disease and Relapse in Central Nervous System Disorders. Innovative Medicine Initiative. URL: http://www.imi.europa.eu/projects-results/project-factsheets/radar-cns [accessed 2021-09-10]
57. Data integration centres. Medical Informatics Initiative Germany. URL: https://www.medizininformatik-initiative.de/en/consortia/data-integration-centres [accessed 2021-09-16]
58. Graphite Health. 2021. URL: https://www.graphitehealth.io/ [accessed 2021-10-11]
59. Sentinel Initiative. URL: https://www.sentinelinitiative.org/ [accessed 2021-09-16]
60. Leijten FRM, Struckmann V, van Ginneken E, Czypionka T, Kraus M, Reiss M, SELFIE consortium. The SELFIE framework for integrated care for multi-morbidity: Development and description. Health Policy 2018 Jan;122(1):12-22 [FREE Full text] [doi: 10.1016/j.healthpol.2017.06.002] [Medline: 28668222]

Abbreviations

CDER: Center for Drug Evaluation and Research
CDRH: Center for Devices and Radiological Health
COPD: chronic obstructive pulmonary disease
DiMe: Digital Medicine Society
EHR: electronic health record
FDA: Food and Drug Administration
RADAR: Remote Assessment of Disease and Relapse
RADAR-CNS: RADAR in Central Nervous System Disorders
SMART on FHIR: Substitutable Medical Applications, Reusable Technologies on Fast Healthcare Interoperability Resources

Edited by G Eysenbach; submitted 26.10.21; this is a non–peer-reviewed article;accepted 27.10.21; published 09.11.21.

Please cite as:
Clay I, Angelopoulos C, Bailey AL, Blocker A, Carini S, Carvajal R, Drummond D, McManus KF, Oakley-Girvan I, Patel KB, Szepietowski P, Goldsack JC
Sensor Data Integration: A New Cross-Industry Collaboration to Articulate Value, Define Needs, and Advance a Framework for Best Practices
J Med Internet Res 2021;23(11):e34493
URL: https://www.jmir.org/2021/11/e34493
doi:10.2196/34493
PMID:34751656

Original Paper

Exploratory Data Mining Techniques (Decision Tree Models) for Examining the Impact of Internet-Based Cognitive Behavioral Therapy for Tinnitus: Machine Learning Approach

Hansapani Rodrigo[1,2], PhD; Eldré W Beukes[2,3,4], PhD; Gerhard Andersson[5,6], PhD; Vinaya Manchaiah[2,3,7,8], MBA, AUD, PhD

[1]School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, Edinburgh, TX, United States

[2]Virtual Hearing Lab, Beaumont, TX, United States

[3]Department of Speech and Hearing Sciences, Lamar University, Beaumont, TX, United States

[4]Department of Vision and Hearing Sciences, School of Psychology and Sport Science, Anglia Ruskin University, Cambridge, United Kingdom

[5]Department of Behavioral Sciences and Learning, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[6]Department of Clinical Neuroscience, Division of Psychiatry, Karolinska Institute, Stockholm, Sweden

[7]Department of Speech and Hearing, School of Allied Health Sciences, Manipal Academy of Higher Education, Karnataka, India

[8]Department of Speech-Language Pathology and Audiology, University of Pretoria, Gauteng, South Africa

Corresponding Author:
Hansapani Rodrigo, PhD
School of Mathematical and Statistical Sciences
University of Texas Rio Grande Valley
1201 W University Drive
Edinburgh, TX, 78539
United States
Phone: 1 9566652313
Email: hansapani.rodrigo@utrgv.edu

Abstract

Background: There is huge variability in the way that individuals with tinnitus respond to interventions. These experiential variations, together with a range of associated etiologies, contribute to tinnitus being a highly heterogeneous condition. Despite this heterogeneity, a “one size fits all” approach is taken when making management recommendations. Although there are various management approaches, not all are equally effective. Psychological approaches such as cognitive behavioral therapy have the most evidence base. Managing tinnitus is challenging due to the significant variations in tinnitus experiences and treatment successes. Tailored interventions based on individual tinnitus profiles may improve outcomes. Predictive models of treatment success are, however, lacking.

Objective: This study aimed to use exploratory data mining techniques (ie, decision tree models) to identify the variables associated with the treatment success of internet-based cognitive behavioral therapy (ICBT) for tinnitus.

Methods: Individuals (N=228) who underwent ICBT in 3 separate clinical trials were included in this analysis. The primary outcome variable was a reduction of 13 points in tinnitus severity, which was measured by using the Tinnitus Functional Index following the intervention. The predictor variables included demographic characteristics, tinnitus and hearing-related variables, and clinical factors (ie, anxiety, depression, insomnia, hyperacusis, hearing disability, cognitive function, and life satisfaction). Analyses were undertaken by using various exploratory machine learning algorithms to identify the most influencing variables. In total, 6 decision tree models were implemented, namely the classification and regression tree (CART), C5.0, GB, XGBoost, AdaBoost algorithm and random forest models. The Shapley additive explanations framework was applied to the two optimal decision tree models to determine relative predictor importance.

Results: Among the six decision tree models, the CART (accuracy: mean 70.7%, SD 2.4%; sensitivity: mean 74%, SD 5.5%; specificity: mean 64%, SD 3.7%; area under the receiver operating characteristic curve [AUC]: mean 0.69, SD 0.001) and gradient boosting (accuracy: mean 71.8%, SD 1.5%; sensitivity: mean 78.3%, SD 2.8%; specificity: 58.7%, SD 4.2%; AUC: mean 0.68, SD 0.02) models were found to be the best predictive models. Although the other models had acceptable accuracy (range 56.3%-66.7%) and sensitivity (range 68.6%-77.9%), they all had relatively weak specificity (range 31.1%-50%) and AUCs (range

0.52-0.62). A higher education level was the most influencing factor for ICBT outcomes. The CART decision tree model identified 3 participant groups who had at least an 85% success probability following the undertaking of ICBT.

Conclusions: Decision tree models, especially the CART and gradient boosting models, appeared to be promising in predicting ICBT outcomes. Their predictive power may be improved by using larger sample sizes and including a wider range of predictive factors in future studies.

(J Med Internet Res 2021;23(11):e28999) doi:10.2196/28999

KEYWORDS

tinnitus; internet interventions; digital therapeutics; cognitive behavioral therapy; artificial intelligence; machine learning; data mining; decision tree; random forest

Introduction

Background

Tinnitus is the perception of a sound in the ears or head in the absence of a corresponding external sound source. It is very prevalent; it has been estimated that 10% to 15% of the adult population experiences tinnitus [1]. Various conditions are associated with developing tinnitus, such as ear disorders [2], exposure to loud noise, the presence of hearing loss, and increasing age [3]. Tinnitus experiences are highly heterogeneous in terms of how it manifests (eg, the types of sounds experienced, how individuals react to their sounds, the associated comorbidities, etc) and how individuals with tinnitus respond to treatment [4]. Although a majority of those with tinnitus are not bothered by their tinnitus, a significant number of people with tinnitus experience distressing tinnitus that affects their quality of life [5]. Nevertheless, tinnitus can affect people in different ways; the most common complaints include annoyance, irritability, fatigue, stress, sleep problems, and trouble concentrating [6]. Moreover, distressing tinnitus is often associated with an increased risk of anxiety and depression [7,8]. Various management strategies are used to help persons with tinnitus, including sound therapy (eg, hearing aids and masking); informational counseling for aiding people's understanding of tinnitus; and psychological approaches that address unhelpful thought patterns and reactions to tinnitus, such as cognitive behavioral therapy (CBT). Of these, CBT has the highest level of research evidence on reducing tinnitus distress [9,10].

Although the use of CBT is recommended in many tinnitus practice guidelines [11], it is seldomly provided. This is partly due to a lack of trained professionals who can offer CBT for tinnitus in an in-person format. To overcome this barrier, internet-based CBT (ICBT) was developed in the late 1990s [12]. In ICBT, the treatment strategies are offered to individuals with tinnitus as self-help materials that are provided over the internet together with professional guidance [13]. The feasibility and efficacy of such an approach have been demonstrated among several populations in Sweden, Germany, Australia, the United Kingdom [14], and, more recently, the United States [15,16]. In general, studies have shown that nearly 50% to 60% of those who undergo ICBT will experience a clinically significant reduction in tinnitus distress [17,18]. To date, no strong predictors of ICBT outcomes have been identified to indicate who is likely to benefit from ICBT interventions. The predictors of outcomes that were identified when examining the long-term (1 year) outcomes of ICBT in the United Kingdom were higher baseline tinnitus severity, more engagement with the ICBT program (ie, more modules opened), and higher self-reported satisfaction with the intervention [18]. To further explore predictors of outcomes, various univariate and multivariate (ie, logistic and linear) regression models were applied to a combined data set of multiple ICBT studies [19]. These linear and logistic regression models identified education level (linear regression: P=.01; logistic regression: P<.001) and baseline tinnitus severity (linear regression: P<.001; logistic regression: P<.001) to be significant predictor variables that contribute to reductions in tinnitus severity following an ICBT intervention. As per the linear regression model, participants who had received disability allowance showed a 25.30-point less (95% CI −46.35 to −4.24) Tinnitus Functional Index (TFI) score reduction when compared to those who did not have to work less due to tinnitus after adjusting for baseline tinnitus severity and participants' education levels.Although many other predictors, including age and tinnitus duration [19], were not identified to be significant under these linear models, these variables might have a nonlinear association with the response.

In the last 2 decades, various artificial intelligence and machine learning techniques have been developed and applied to hearing health data. Such approaches have mainly been used for disease profiling, although some studies have focused on the prediction of treatment outcomes [20-24]. It is noteworthy that the intervention trials in audiology and tinnitus research usually involve a few hundred participants and the collection of generally extensive data regarding demographic characteristics and clinical variables. Such a data set with many predictor variables may be best handled by exploratory data mining techniques, such as the use of tree-based models (eg, the random forest [RF] model). Such models tend to perform well even in the presence of multicollinearity among a large number of predictor variables, as these models decorrelate the variables [25]. For example, a recent study that examined various machine learning algorithms for predicting CBT treatment outcomes in the tinnitus population suggested that gradient boosted trees (area under the receiver operating characteristic curve [AUC]=0.89) have the best predictive power [21]. This study found that subjectively perceived tinnitus-related impairment, depression, sleep problems, physical health–related impairments in quality of life, the time spent on completing questionnaires, and educational level highly contributed to the model's predictions. However, no previous studies have examined the

application of artificial intelligence and machine learning techniques to ICBT outcomes in tinnitus research.

Objectives

To further explore outcome predictors for ICBT, this study aimed to examine the applications of various exploratory data mining techniques based on decision tree models. In particular, we wanted to (1) investigate which types of decision tree models were the most applicable to ICBT outcome prediction (ie, models with the best accuracy and predictive power) and (2) identify the most relevant predictive factors of ICBT outcomes by using the most appropriate decision tree models.

Methods

Study Design and Ethical Considerations

We included 228 participants who previously underwent ICBT for tinnitus and whose data were collected as a part of 3 separate ICBT trials [17,18,26] that were conducted from 2016 to 2018. This study was a secondary analysis of these ICBT intervention studies. Ethical clearance was obtained from the Faculty of Science and Technology Research Ethics Panel of Anglia Ruskin University (reference numbers: FST/FREP/14/478 and FST/FREP/14/478) and the East of England–Cambridge South Research Ethics Committee (reference number: 16/EE/0148) and Health Research Authority (Integrated Research Application System project ID: 195565).

Participant Characteristics

A heterogenous sample of individuals with tinnitus was obtained; thus, the sample represented typical tinnitus populations, as seen in Multimedia Appendix 1. The average age was 55.14 years (SD 12.92 years), and 98 out of the 228 (43%) participants were females. The majority of participants (154/228, 67.5%) had long-standing tinnitus with a mean duration of 17.68 years (SD 19.42 years). Of the 228 participants, 59 (25.9%) had completed high school education, 61 (26.8%) had an undergraduate degree, and only 30 (13.2%) had a postgraduate degree. Approximately 48% (109/228, 47.8%) of the participants experienced tinnitus in both ears, 26.8% (61/228) of them experienced tinnitus in 1 ear, and the others reported experiencing tinnitus in their head or in other locations. The majority (159/228, 69.7%) of participants did not wear hearing aids, and 25.4% (58/228) of them had sought tinnitus treatment previously.

Intervention

The study participants completed an 8-week ICBT intervention that was presented in a self-help format [13,27]. The intervention was administered by using a secure e-platform [28,29]. During this 8-week period, participants were presented with 2 to 3 learning modules that contained various elements of CBT that were specifically adapted for tinnitus, including applied relaxation, cognitive restructuring, and imagery. The digital materials were presented by using text, images, and videos. In addition, various exercises were presented in these learning modules to improve engagement.

Data Collection

The baseline data collection included an extensive questionnaire that focused on demographics and tinnitus-related and treatment-related information. Outcome data were gathered by using standardized primary and secondary self-reported questionnaires, which were administered before (baseline), during (weekly), and after the intervention. The primary outcome was a change in tinnitus severity, as measured by the TFI [30]. The secondary outcome measures included the Insomnia Severity Index [31] (a measure of insomnia), the Generalized Anxiety Disorder-7 [32] (a measure of anxiety); the Patient Health Questionnaire-9 [33] (a measure of depressive symptoms); the Hearing Handicap Inventory for Adults Screening version [34] (a measure of self-reported hearing disability); the Hyperacusis Questionnaire [35], which was used to assess the presence hyperacusis (ie, reduced tolerance to everyday sounds); the Cognitive Failures Questionnaire [36], which was used to assess cognitive functions; and the Satisfaction with Life Scales [37], which were used to assess global life satisfaction.

Data Analyses

Variables

The primary outcome variable (the dependent variable) in this study was a change in tinnitus severity. A 13-point reduction in TFI scores following the ICBT intervention was regarded as a clinically significant (successful) treatment outcome [30]. Significant differences in scores were assessed by using paired sample *t* tests. All tests were two-tailed, and significance was set to P=.05. There were 33 predictor variables selected, as outlined in Multimedia Appendix 2. These included the following:

- 7 demographic variables (ie, age, gender, education level, employment type, noise exposure, the presence of psychological conditions, and tinnitus that affects the ability to work)
- 15 tinnitus and hearing-related variables (ie, baseline tinnitus severity, tinnitus duration, how often tinnitus is heard, tinnitus location, 9 different types of tinnitus, tinnitus in which multiple tones are heard, and the presence of hearing loss)
- 4 treatment-related variables (ie, past treatment sought, tinnitus maskability, hearing aid use, and medication use)
- 7 clinical factors (ie, anxiety, depression, insomnia, hyperacusis, hearing disability, cognitive functions, and life satisfaction).

Decision Tree Models (Classifiers)

The data analysis focused on decision tree–based models, as they play an essential role in exploratory data mining and facilitate human decision-making by providing decision rules [38]. Despite their simplicity, decision trees usually exhibit high variance in their predictions and are not consistently robust. Given these issues, their powerful counterparts, such as the RF [39], gradient boosting (GB) [40], and extreme GB (XGBoost) [41], models were selected. For comparison, 6 decision tree models were used, namely the classification and regression tree (CART) [42], C5.0 [43], GB, XGBoost, AdaBoost algorithm

[44], and RF models. As the CART, C5.0, and RF decision tree models involve stratifying or segmenting the predictor space into a number of nonoverlapping regions [38], recursive binary splitting for classification via the Gini index was performed [45]. Many of these decision tree types have been applied to audiological data, and they were found to provide good results in previous studies [20-24].

Data Analysis Steps

Summary of Data Analysis Steps

The analyses were performed in 4 stages. First, the data were split into training and testing data, and the classifier models were trained on a data set before testing. Second, the six classifiers were applied to the test data to identify the most suitable models based on their performance evaluation. Third, the two best models were used to determine the predictors of ICBT outcomes. Fourth, the optimal CART decision tree model was used to identify the participants who were more (or less) likely to benefit from ICBT treatment. The steps are described in more detail below.

Step 1: Classifier Training

Prior to applying the decision tree classifiers, the entire data set was divided into the training (183/228, 80.3%) and testing (45/228, 19.7%) data sets. The training data set was used to develop the corresponding data mining model, while the testing data set was used to evaluate the model predictions. As the training data set was relatively small (n=183), a repeated 3-fold cross-validation was performed. This was not done for the CART model, for which the full training data set was used for model training. With this approach, each fold was given a chance to act as their own validation set to minimize the propensity of model overfitting. In total, 10 different models were created with several random initializations for each data mining method. Hyperparameter tuning for each of these decision models was performed, as required. For instance, when training the RF models, we explored a range of different numbers of predictors for the splitting at each tree node and their impact on the models' performance.

Step 2: Classifier Performance Evaluation

The trained models were evaluated (ie, by using the testing data set) in terms of their mean predictive accuracy, sensitivity (true positive rate), specificity (true negative rate), and AUCs. These were presented as means and SDs and based on the 10 replicated models for each data mining technique. The AUC is used as a measurement of model discrimination power. The optimal decision tree models were selected based on having the highest AUC value. In general, models with an AUC of 0.5 have no discriminatory power, models with an AUC of 0.7 to 0.8 are considered acceptable, models with an AUC of 0.8 to 0.9 are considered excellent, and models with an AUC of >0.9 are deemed to have outstanding discriminatory power (ie, the ability to identify patients with and without a disease or condition based on a new set of data).

Step 3: Predictors of ICBT Outcomes

Decision tree–based classifiers provide insights on different participant groups who show promising results following ICBT. After identifying the two most optimal models, the model-agnostic posthoc framework Shapley additive explanations (SHAP) was used for analyzing ICBT outcome predictors [46,47]. This framework facilitates model interpretations and identifies the most influential factors that result in successful ICBT outcomes (ie, a reduction in TFI scores following the ICBT intervention). SHAP measure the impact of variables and take into account variables' interactions with other variables. SHAP values indicate the importance of a feature and are calculated by comparing model predictions that account and do not account for a given feature. However, since the order in which a model sees features can affect its predictions, this comparison is done in every possible order, so that the features are compared in a fair manner.

Step 4: Identification of Participants Who Are Most Likely to Benefit From ICBT

The CART decision tree model was used to identify the participants who were the most (or least) likely to benefit from ICBT. In training, a minimum split of 20 and a max depth of 10 were used as the control parameters for the CART decision tree models. Tree pruning was conducted to reduce the overfitting in the CART decision tree models, although the best decision tree model remains the same even after pruning.

The data analysis was performed with R version 4.0.3 (R Foundation for Statistical Computing) software. The code is available in the GitHub repository for this study [48]. The data can be made available upon reasonable request.

Results

ICBT Effects

Undertaking ICBT significantly reduced tinnitus severity scores (t_{227}=16.37; P<.001) from a mean baseline severity score of 57.93 (SD 19.17) to a mean post-ICBT severity score of 34.22 (SD 22.78), as measured by the TFI. A clinically significant 13-point change in TFI scores was achieved by 150 of the 228 participants (65.8%) after the intervention.

Decision Tree Model Performance Evaluations

Table 1 contains the model evaluation information of all 6 decision tree classifiers that were based on the test data. Following training via the 3-fold cross-validation method, the mean accuracies of the six decision tree classifiers ranged between a minimum of 56.3% (the C5.0 model) to a maximum of 71.8% (the GB model). Model predictions showed variations in their sensitivity (range 68.6%-78.3%), specificity (range 31.1%-64%), and AUC values (range 0.52-0.69).

Table 1. Decision tree model evaluations.

Classification model	Accuracy (%), mean (SD)	Sensitivity (%; true positive rate), mean (SD)	Specificity (%; true negative rate), mean (SD)	AUC[a], mean (SD)
Classification and regression decision tree	70.7 (2.4)	74 (5.5)	64 (3.7)	0.69 (0.001)
C5.0	56.3 (1.1)	68.6 (1.9)	31.1 (6.3)	0.52 (0.001)
Gradient boosting	71.8 (1.5)	78.3 (2.8)	58.7 (4.2)	0.68 (0.02)
Extreme gradient boosting	65 (4.1)	77.9 (8.7)	39.2 (6.6)	0.62 (0.08)
AdaBoost algorithm	63.6 (3.2)	73.3 (5.2)	44 (7.8)	0.58 (0.05)
Random forest	66.7 (3)	75 (6.1)	50 (7.2)	0.60 (0.01)

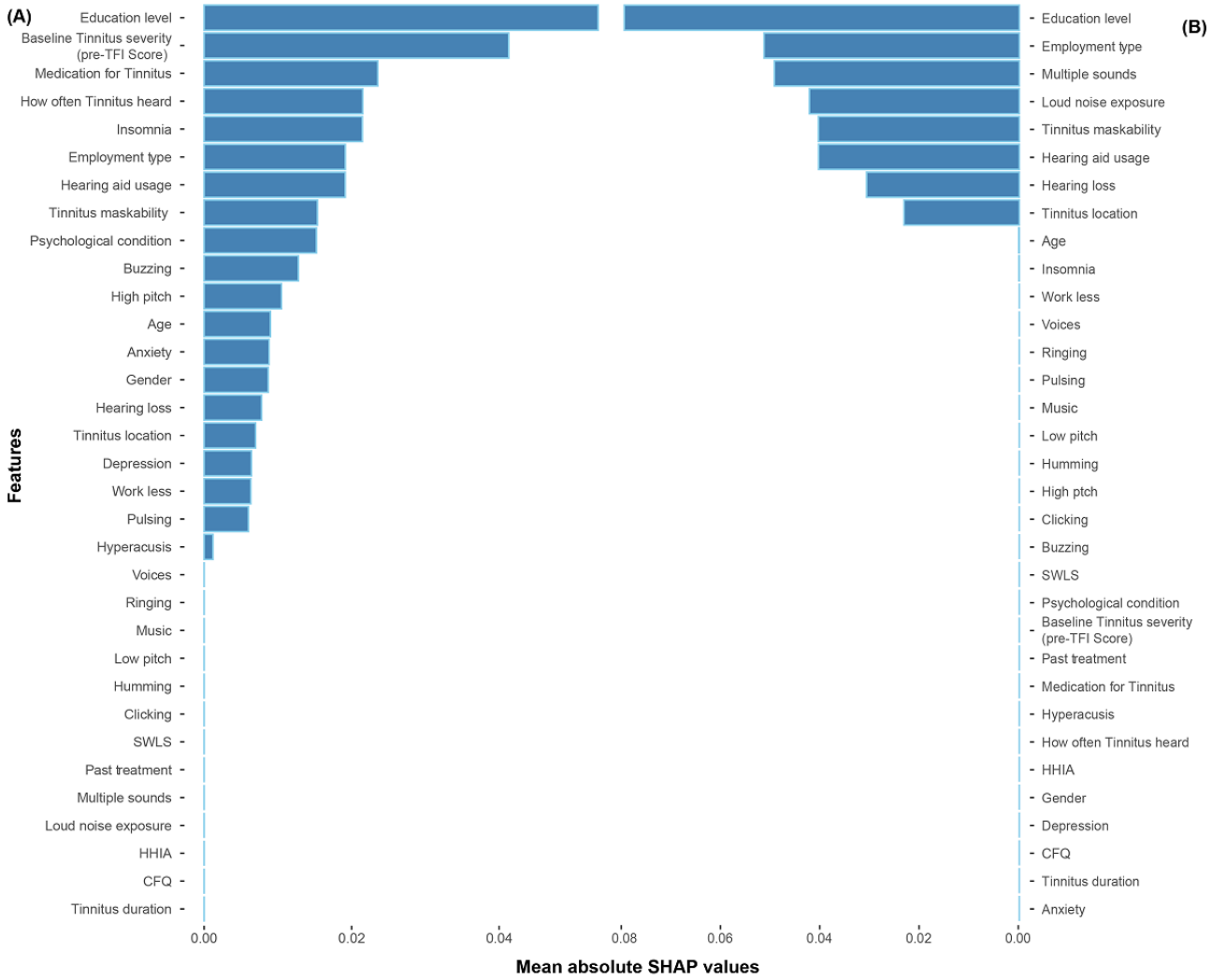
[a]AUC: area under the receiver operating characteristic curve.

None of the six models were considered robust, as their AUC values were below 0.80. As the CART and GB classifiers were found to be superior compared to the other four models when considering all evaluation measurements as a whole (accuracy, sensitivity, specificity, and the AUC), these two models were further examined.

Feature Importance

The SHAP framework was applied to the CART and GB classifiers to estimate each predictor variable's importance in predicting ICBT outcomes (Figure 1). Variables with larger SHAP values are relatively more important in terms of their contributions (ie, feature contribution) to a model prediction. Education level (average SHAP values: GB model=0.053; CART model=0.079) was identified as the most important influencing factor in both models. Although they were not ranked in the same order, the other features that ranked within the top 10 features for both models were employment type (average SHAP values: GB model=0.019; CART model=0.051), hearing aid usage (average SHAP values: GB model=0.019; CART model=0.040), and tinnitus maskability (average SHAP value: GB model=0.015; CART model=0.040). The differences between these models were that the GB model ranked baseline tinnitus severity (average SHAP value=0.041), how often tinnitus is heard (average SHAP value=0.022), insomnia (average SHAP value=0.021), the use of medication for tinnitus (average SHAP value=0.024), and the presence of a psychological condition (average SHAP value=0.015) among the top 10 features, whereas the CART model ranked the presence of multiple sounds (average SHAP value=0.049), loud noise exposure (average SHAP value=0.042), and tinnitus location (average SHAP value=0.023) as key features.

Figure 1. Feature importance based on the mean absolute SHAP values from (A) the best gradient model and (B) from the best classification and regression tree decision tree model. These SHAP values represent the absolute change in log odds. Relatively higher importance is indicated with larger SHAP values. CFQ: Cognitive Failures Questionnaire; HHIA: Hearing Handicap Inventory for Adults; SHAP: Shapley additive explanations; SWLS: Satisfaction with Life Scales; TFI: Tinnitus Functional Index.



Figures 2 and 3 present the effect that each feature category had on the outcome variable, as decided by the best GB and CART decision tree models. Features' impacts on the two classes are presented in 2 separate plots for each feature ("1" indicates the effect on the successful treatment class and "0" indicates the effect on the unsuccessful treatment class). Positive SHAP values in each successful treatment group indicate a higher log odd of achieving a 13-point or more tinnitus severity score reduction (ie, on the TFI for a given category) for a feature and vice versa. This log odd is relative to the training set average.

Figures 2 and 3 both depict positive SHAP values for the participants who had a vocational training degree or a master's degree (or above), those with higher levels of education (postgraduate degrees), and those who were using a hearing aid in only 1 ear. As per the GB model, this reduction was more likely to occur for participants with insomnia (scores of 14 or less on the Insomnia Severity Index), a psychological condition, and tinnitus that can be described as a buzzing sound and for participants who had median baseline tinnitus severity scores (ie, pre-TFI assessment) of >55.2.

Figure 2. The best gradient boosting model–based feature effects. Each graph represents a feature and their corresponding SHAP values. Plots labeled with "1" illustrate the impact that each feature has on achieving a successful treatment outcome (a 13-point or more reduction in TFI score). CFQ: Cognitive Failures Questionnaire; HHIA: Hearing Handicap Inventory for Adults; SHAP: Shapley additive explanations; SWLS: Satisfaction with Life Scales; TFI: Tinnitus Functional Index.

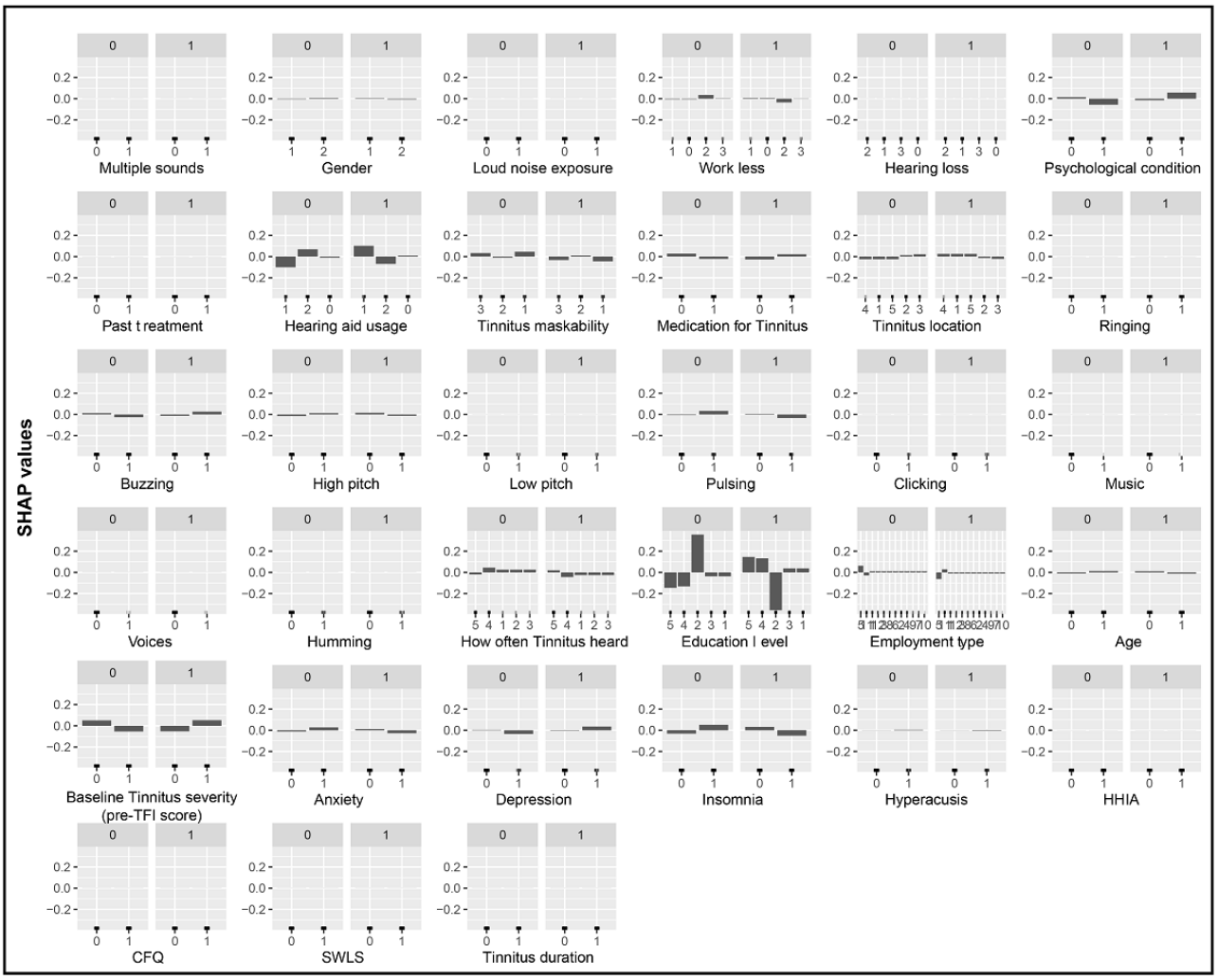
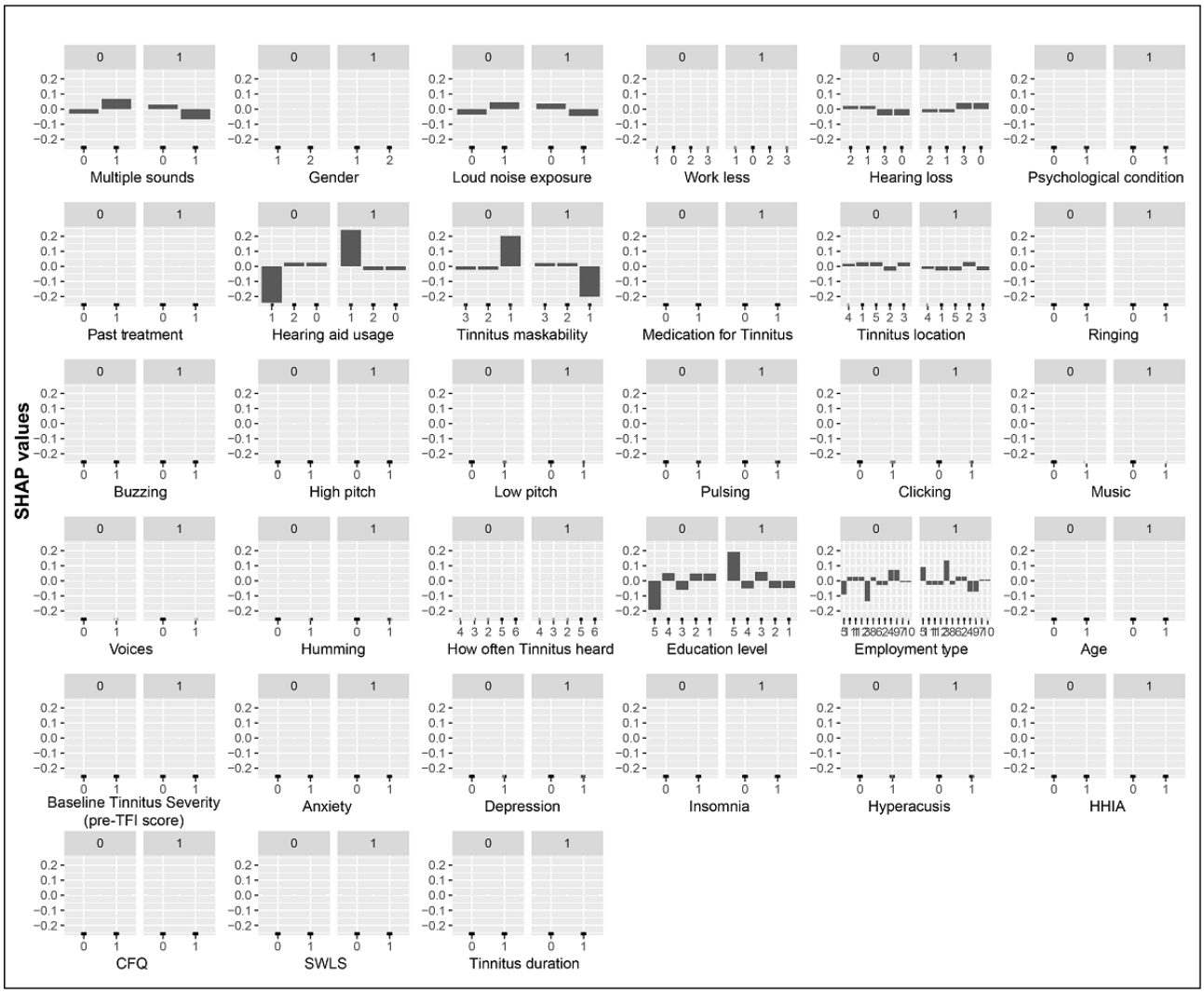


Figure 3. The best classification and regression tree decision tree model feature effects. Each graph represents a feature vs. corresponding SHAP value. Plots with "1" represent the effect that each feature has on achieving a successful treatment outcome. CFQ: Cognitive Failures Questionnaire; HHIA: Hearing Handicap Inventory for Adults; SHAP: Shapley additive explanations; SWLS: Satisfaction with Life Scales; TFI: Tinnitus Functional Index.

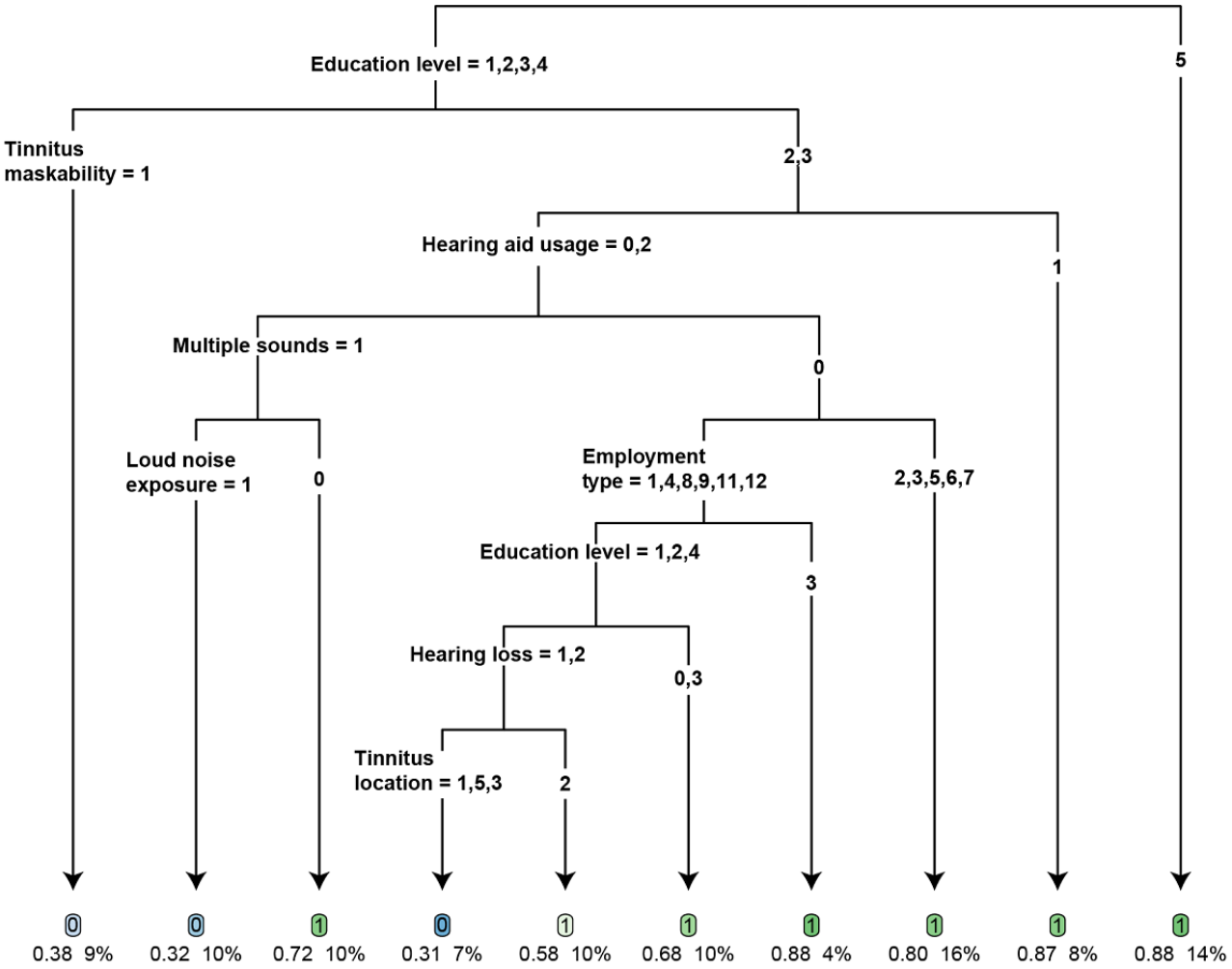


Identification of Participants Who Are Likely to Benefit From ICBT

Figure 4 presents the final decision tree model, which has 10 nodes. A detailed explanation is provided in Multimedia Appendix 3. In this model, homogenous groups were formed by creating binary splits at each node. The decision nodes represent the treatment groups (either 0 or 1) that were the most likely to achieve tinnitus severity reduction, and the decisions in the tree were based on the characteristics of each group. These characteristics were represented by each branch (based on their corresponding feature values) of the tree. This model showed that higher education level, tinnitus maskability, hearing aid usage, the presence of multiple tinnitus sounds, loud noise exposure, employment type, the presence of hearing loss, and tinnitus location were important factors for determining treatment outcomes. The following participant groups had at least an 85% chance of achieving a TFI score reduction of 13 points or more:

- Participants with postgraduate education (master's degree or higher)
- Participants with an education level other than a master's level of education and poor tinnitus maskability (or only partial tinnitus maskability) and those who wore a hearing aid in 1 ear
- Participants with no tinnitus maskability or only partial tinnitus maskability; those who did not wear a hearing aid or used hearing aids bilaterally; those who did not hear multiple tinnitus sounds; and those with an occupation that could described as *professional*, *technical*, *skills based* (ie, skilled tradesman), *service related*, or *medical*

Figure 4. The best classification and regression tree decision tree model. The fitted tree has 10 terminal nodes (denote the decision criteria). The variable categories (please refer to Multimedia Appendix 2 for variable category labels) corresponding to each split are given at the top of each branch. Each terminal node contains the predicted treatment class ("1" indicates the successful treatment class and "0" indicates the unsuccessful treatment class), the percentage of subjects with successful treatment outcomes, and the percentage of participants with the given characteristics in the training set.



Discussion

The aim of this study was to explore predictors of outcomes in ICBT for tinnitus by applying 6 types of decision tree models to a combined data set of participants from 3 clinical trials. The key findings are discussed below.

Best Decision Tree Models

In this study, we applied 6 different decision tree models to ICBT data. Although none of the six models attained an excellent or outstanding status, the CART and GB models' discriminative power can be considered to be satisfactory, given the moderate sample size with just 33 predictive factors (features). This is consistent with a recent study that used 10 decision tree models to predict the outcomes of CBT for patients with tinnitus (N=1416) and found that the GB model with 26 predictive factors had the best predictive power; it had an AUC value of 0.89 [21].

Further work however is needed in this area to determine which models and how many vital factors may result in an optimal predictive model. A larger sample size would likely improve the results. However, we are not sure if just adding factors would be helpful. This is because Niemann et al [21] included 205 factors in their analysis, and of these, only 26 were helpful in achieving optimal results. Moreover, of the 26 factors, only a handful had the most considerable effect. For instance, only 1 factor (ie, tinnitus impairment in terms of loudness, frequency, and distress) resulted in an AUC of 0.79, only 3 features resulted in an AUC of 0.85, and only 8 factors resulted in an AUC of 0.85. These results indicate that including key factors with high predictive power may be a better approach than just adding all of the possible factors.

Predictors of ICBT Outcomes

Among the best decision tree models (ie, the CART and GB models), various factors were found to be critical predictors of ICBT outcomes. These included demographics (ie, education level, employment type, and the presence of a psychological condition), tinnitus and hearing-related factors (ie, baseline tinnitus severity, tinnitus location, how often tinnitus is heard, a buzzing type of tinnitus, tinnitus maskability, and hearing loss type), treatment-related factors (ie, hearing aid usage), and clinical factors (ie, insomnia). However, education level was the most notable predictor among these.

Participants who had a master's degree or above had an 88% chance of achieving a successful outcome. This is

understandable, as the ability to read, understand, and follow instructions is key for undergoing self-help interventions. However, it is likely that the way in which the materials were written may also have played a role. For instance, UK ICBT materials are written at a ninth-grade reading level [49], which may require higher literacy skills. However, these materials have been rewritten at a sixth-grade reading level and below [49] to ensure accessibility for those with lower education levels. More and more people, including those with lower education, are using the internet and participating in internet-based treatments, particularly due to the constraints that have been placed on health care during the COVID-19 pandemic [50-52]. In our sample, over 85% (198/228, 86.8%) of the participants had a below–master-level education. This highlights the need for making ICBT more accessible to increase the chances of achieving improved outcomes for those with lower education levels.

Baseline tinnitus severity was found to be another critical factor for predicting ICBT outcomes in the GB model. Our previous studies on 1-year outcomes [53] and our previous application of univariate and multivariate analyses to this study's sample [19] identified baseline tinnitus severity as a critical predictive variable. In the Niemann et al [21] study, tinnitus loudness, frequency, and distress, which were measured by using a visual analog scale, were found to be key predictive factors. Further, tinnitus distress, which was measured by using the German version of the Tinnitus Questionnaire (a tool that is comparable to the TFI in this study), was not found to be the key predictive factor. However, based on both clinical experiences and findings from many previous studies, baseline tinnitus severity is an important factor for determining treatment outcomes. The clinical factors depression and anxiety were among the key predictive factors in the GB model. A recent clinical trial conducted by Beukes et al [54], as well as the Niemann et al [21] study, determined that those with high levels of depression had a better chance of achieving success.

Although various other tinnitus and hearing-related variables could have played a role in determining the outcomes of ICBT, the predictive power of our models was relatively low based on the AUC values. Nevertheless, it would be useful for hearing health care professionals to examine these factors when deciding on the candidacy of self-help psychological interventions such as ICBT. Moreover, it would be useful for future studies to examine any additional factors (eg, health literacy) that may have a bearing on ICBT outcomes.

Study Limitations and Future Directions

Although this study is among the first to apply data mining models to ICBT data, it has several limitations. The sample size was limited, and this may have contributed to the low predictive accuracies of the models. The exploratory decision tree models worked better when including a large number of predictive factors. In this study, we only included 33 predictive factors in our models, and this may have limited the performance of our models. Further, we may have missed some important factors (eg, health literacy) that have a bearing on ICBT outcomes.

The inclusion and exclusion criteria that were used in the three trials from which our data were generated may have resulted in a sample with high tinnitus severity levels that may not be representative of the general tinnitus population. This may have also contributed to our limited key findings. Future studies could include more extensive samples of heterogeneous patients with tinnitus as well as all of the possible predictive factors that could help with improving our models' predictive power. Moreover, developing nonlinear classifiers with artificial neural networks and support vector machines could help with achieving higher prediction accuracies and should be examined in future studies.

In conclusion, tree models, especially the CART and GB models, appear to be promising in predicting ICBT outcomes. Future studies should be undertaken with larger sample sizes and include a more comprehensive range of predictive factors to improve their models' predictive power.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the study's participants.
[DOCX File , 28 KB - jmir_v23i11e28999_app1.docx]

Multimedia Appendix 2
Predictor variables.
[DOCX File , 29 KB - jmir_v23i11e28999_app2.docx]

Multimedia Appendix 3
Detailed explanation of classification and regression tree model.
[DOCX File , 25 KB - jmir_v23i11e28999_app3.docx]

References

1. McCormack A, Edmondson-Jones M, Somerset S, Hall D. A systematic review of the reporting of tinnitus prevalence and severity. Hear Res 2016 Jul;337:70-79 [FREE Full text] [doi: 10.1016/j.heares.2016.05.009] [Medline: 27246985]

2. Kostev K, Alymova S, Kössl M, Jacob L. Risk factors for tinnitus in 37,692 patients followed in general practices in Germany. Otol neurotol 2019 Apr;40(4):436-440. [doi: 10.1097/mao.0000000000002161] [Medline: 30870350]
3. Kim HJ, Lee HJ, An SY, Sim S, Park B, Kim SW, et al. Analysis of the prevalence and associated risk factors of tinnitus in adults. PLoS One 2015 May 28;10(5):e0127578 [FREE Full text] [doi: 10.1371/journal.pone.0127578] [Medline: 26020239]
4. Beukes EW, Manchaiah V, Allen PM, Andersson G, Baguley DM. Exploring tinnitus heterogeneity. Prog Brain Res 2021;260:79-99. [doi: 10.1016/bs.pbr.2020.05.022] [Medline: 33637233]
5. Manchaiah V, Beukes EW, Granberg S, Durisala N, Baguley DM, Allen PM, et al. Problems and life effects experienced by tinnitus research study volunteers: An exploratory study using the ICF classification. J Am Acad Audiol 2018;29(10):936-947. [doi: 10.3766/jaaa.17094] [Medline: 30479266]
6. Beukes EW, Manchaiah V, Andersson G, Allen PM, Terlizzi PM, Baguley DM. Situationally influenced tinnitus coping strategies: a mixed methods approach. Disabil Rehabil 2018 Dec;40(24):2884-2894. [doi: 10.1080/09638288.2017.1362708] [Medline: 28793790]
7. Salazar JW, Meisel K, Smith ER, Quiggle A, McCoy DB, Amans MR. Depression in patients with tinnitus: A systematic review. Otolaryngol Head Neck Surg 2019 Jul;161(1):28-35. [doi: 10.1177/0194599819835178] [Medline: 30909841]
8. Trevis KJ, McLachlan NM, Wilson SJ. A systematic review and meta-analysis of psychological functioning in chronic tinnitus. Clin Psychol Rev 2018 Mar;60:62-86. [doi: 10.1016/j.cpr.2017.12.006] [Medline: 29366511]
9. Fuller T, Cima R, Langguth B, Mazurek B, Vlaeyen JW, Hoare DJ. Cognitive behavioural therapy for tinnitus. Cochrane Database Syst Rev 2020 Jan 08;1(1):CD012614 [FREE Full text] [doi: 10.1002/14651858.CD012614.pub2] [Medline: 31912887]
10. Landry EC, Sandoval XCR, Simeone CN, Tidball G, Lea J, Westerberg BD. Systematic review and network meta-analysis of cognitive and/or behavioral therapies (CBT) for tinnitus. Otol Neurotol 2020 Feb;41(2):153-166. [doi: 10.1097/MAO.0000000000002472] [Medline: 31743297]
11. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham Jr ER, et al. Clinical practice guideline: tinnitus executive summary. Otolaryngol Head Neck Surg 2014 Oct;151(4):533-541. [doi: 10.1177/0194599814547475] [Medline: 25274374]
12. Andersson G, Strömgren T, Ström L, Lyttkens L. Randomized controlled trial of internet-based cognitive behavior therapy for distress associated with tinnitus. Psychosom Med 2002;64(5):810-816. [doi: 10.1097/01.psy.0000031577.42041.f8] [Medline: 12271112]
13. Beukes EW, Vlaescu G, Manchaiah V, Baguley DM, Allen PM, Kaldo V, et al. Development and technical functionality of an internet-based intervention for tinnitus in the UK. Internet Interv 2016 Aug 14;6:6-15 [FREE Full text] [doi: 10.1016/j.invent.2016.08.002] [Medline: 30135809]
14. Beukes EW, Manchaiah V, Allen PM, Baguley DM, Andersson G. Internet-based interventions for adults with hearing loss, tinnitus, and vestibular disorders: A systematic review and meta-analysis. Trends Hear 2019;23:2331216519851749 [FREE Full text] [doi: 10.1177/2331216519851749] [Medline: 31328660]
15. Beukes EW, Andersson G, Fagelson M, Manchaiah V. Audiologist-supported internet-based cognitive behavioral therapy for tinnitus in the United States: A pilot trial. Am J Audiol 2021 Sep 10;30(3):717-729. [doi: 10.1044/2021_AJA-20-00222] [Medline: 34432984]
16. Beukes E, Andersson G, Fagelson M, Manchaiah V. Internet-based audiologist-guided cognitive behavioral therapy tinnitus in the United States: A randomized controlled trial. JMIR Preprints. Preprint posted online on January 30, 2021 [FREE Full text] [doi: 10.2196/preprints.27584]
17. Beukes EW, Baguley DM, Allen PM, Manchaiah V, Andersson G. Audiologist-guided internet-based cognitive behavior therapy for adults with tinnitus in the United Kingdom: A randomized controlled trial. Ear Hear 2018;39(3):423-433. [doi: 10.1097/AUD.0000000000000505] [Medline: 29095725]
18. Beukes EW, Andersson G, Allen PM, Manchaiah V, Baguley DM. Effectiveness of guided internet-based cognitive behavioral therapy vs face-to-face clinical care for treatment of tinnitus: A randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2018 Dec 01;144(12):1126-1133 [FREE Full text] [doi: 10.1001/jamaoto.2018.2238] [Medline: 30286238]
19. Rodrigo H, Beukes EW, Andersson G, Manchaiah V. Internet-based cognitive-behavioural therapy for tinnitus: secondary analysis to examine predictors of outcomes. BMJ Open 2021 Aug 20;11(8):e049384 [FREE Full text] [doi: 10.1136/bmjopen-2021-049384] [Medline: 34417217]
20. Juhola M, Viikki K, Laurikkala J, Auramo Y, Kentala E, Pyykkö I. Application of artificial intelligence in audiology. Scand Audiol Suppl 2001(52):97-99. [doi: 10.1080/010503901300007209] [Medline: 11318498]
21. Niemann U, Boecking B, Brueggemann P, Mebus W, Mazurek B, Spiliopoulou M. Tinnitus-related distress after multimodal treatment can be characterized using a key subset of baseline variables. PLoS One 2020 Jan 30;15(1):e0228037. [doi: 10.1371/journal.pone.0228037] [Medline: 31999776]
22. Rasku J, Pyykkö I, Levo H, Kentala E, Manchaiah V. Disease profiling for computerized peer support of Ménière's disease. JMIR Rehabil Assist Technol 2015 Sep 03;2(2):e9 [FREE Full text] [doi: 10.2196/rehab.4109] [Medline: 28582248]

23. Shoushtarian M, Alizadehsani R, Khosravi A, Acevedo N, McKay CM, Nahavandi S, et al. Objective measurement of tinnitus using functional near-infrared spectroscopy and machine learning. PLoS One 2020 Nov 18;15(11):e0241695. [doi: 10.1371/journal.pone.0241695] [Medline: 33206675]
24. Weininger O, Warnecke A, Lesinski-Schiedat A, Lenarz T, Stolle S. Computational analysis based on audioprofiles: A new possibility for patient stratification in office-based otology. Audiol Res 2019 Nov 05;9(2):230 [FREE Full text] [doi: 10.4081/audiores.2019.230] [Medline: 31728177]
25. Raj S. Effects of multi-collinearity in logistic regression, SVM, random forest(RF). Medium. 2019 Aug 09. URL: https://medium.com/@raj5287/effects-of-multi-collinearity-in-logistic-regression-svm-rf-af6766d91f1b [accessed 2021-09-22]
26. Beukes EW, Allen PM, Manchaiah V, Baguley DM, Andersson G. Internet-based intervention for tinnitus: Outcome of a single-group open trial. J Am Acad Audiol 2017 Apr;28(4):340-351. [doi: 10.3766/jaaa.16055] [Medline: 28418328]
27. Beukes EW, Andersson G, Manchaiah V, Kaldo V. Cognitive Behavioral Therapy for Tinnitus. USA: Plural Publishing; Oct 28, 2020.
28. Manchaiah V, Vlaescu G, Varadaraj S, Aronson EP, Fagelson MA, Munoz MF, et al. Features, functionality, and acceptability of internet-based cognitive behavioral therapy for tinnitus in the United States. Am J Audiol 2020 Sep 03;29(3):476-490 [FREE Full text] [doi: 10.1044/2020_AJA-20-00002] [Medline: 32880499]
29. 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 Oct 03;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
30. Meikle M, Henry J, Griest S, Stewart B, Abrams H, McArdle R, et al. The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. Ear Hear 2012;33(2):153-176. [doi: 10.1097/AUD.0b013e31822f67c0] [Medline: 22156949]
31. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
32. 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]
33. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
34. Newman CW, Weinstein BE, Jacobson GP, Hug GA. Test-retest reliability of the hearing handicap inventory for adults. Ear Hear 1991 Oct;12(5):355-357. [doi: 10.1097/00003446-199110000-00009] [Medline: 1783240]
35. Khalfa S, Dubal S, Veuillet E, Perez-Diaz F, Jouvent R, Collet L. Psychometric normalization of a hyperacusis questionnaire. ORL J Otorhinolaryngol Relat Spec 2002;64(6):436-442. [doi: 10.1159/000067570] [Medline: 12499770]
36. Broadbent DE, Cooper PF, FitzGerald P, Parkes KR. The Cognitive Failures Questionnaire (CFQ) and its correlates. Br J Clin Psychol 1982 Feb;21(1):1-16. [doi: 10.1111/j.2044-8260.1982.tb01421.x] [Medline: 7126941]
37. 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]
38. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. USA: Springer; 2013.
39. Breiman L. Random forests. Machine learning 2001 Oct;45:5-32. [doi: 10.1023/A:1010933404324]
40. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001 Oct;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
41. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. 2016 Aug Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, California, USA p. 785-794. [doi: 10.1145/2939672.2939785]
42. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Boca Raton, Florida, USA: Chapman and Hall/CRC; 1984.
43. Quinlan J. C4.5: Programs for Machine Learning. San Francisco, California, USA: Morgan Kaufmann Publishers; Mar 1993.
44. Gandhi R. Gradient boosting and XGBoost. Hacker Noon. 2018 May 06. URL: https://hackernoon.com/gradient-boosting-and-xgboost-90862daa6c77 [accessed 2021-09-22]
45. Gini C. On the Measure of Concentration with Special Reference to Income and Statistics. Colorado College Publication 1936;208:73-79.
46. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Dec Presented at: The 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, California, USA p. 4768-4777.
47. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
48. Hansapani / AI-Tinnitus. GitHub. URL: https://github.com/Hansapani/AI-Tinnitus [accessed 2021-09-23]

49. Beukes EW, Fagelson M, Aronson EP, Munoz MF, Andersson G, Manchaiah V. Readability following cultural and linguistic adaptations of an internet-based intervention for tinnitus for use in the United States. Am J Audiol 2020 Jun 08;29(2):97-109 [FREE Full text] [doi: 10.1044/2019_AJA-19-00014] [Medline: 32105512]
50. Beukes EW, Baguley DM, Jacquemin L, Lourenco MPCG, Allen PM, Onozuka J, et al. Changes in tinnitus experiences during the COVID-19 pandemic. Front Public Health 2020 Nov 05;8:592878 [FREE Full text] [doi: 10.3389/fpubh.2020.592878] [Medline: 33251179]
51. Beukes EW, Onozuka J, Brazell TP, Manchaiah V. Coping with tinnitus during the COVID-19 pandemic. Am J Audiol 2021 Jun 14;30(2):385-393. [doi: 10.1044/2021_AJA-20-00188] [Medline: 33979227]
52. 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]
53. Beukes EW, Allen PM, Baguley DM, Manchaiah V, Andersson G. Long-term efficacy of audiologist-guided internet-based cognitive behavior therapy for tinnitus. Am J Audiol 2018 Nov 19;27(3S):431-447 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0004] [Medline: 30452747]
54. Beukes EW, Andersson G, Fagelson MA, Manchaiah V. Dismantling internet-based cognitive behavioral therapy for tinnitus. The contribution of applied relaxation: A randomized controlled trial. Internet Interv 2021 May 12;25:100402 [FREE Full text] [doi: 10.1016/j.invent.2021.100402] [Medline: 34040997]

Abbreviations

AUC: area under the receiver operating characteristic curve
CART: classification and regression tree
CBT: cognitive behavioral therapy
GB: gradient boosting
ICBT: internet-based cognitive behavioral therapy
RF: random forest
SHAP: Shapley additive explanations
TFI: Tinnitus Functional Index
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 21.03.21; peer-reviewed by G Liu, JA Benítez-Andrades; comments to author 28.06.21; revised version received 29.08.21; accepted 12.09.21; published 02.11.21.

Please cite as:
Rodrigo H, Beukes EW, Andersson G, Manchaiah V
Exploratory Data Mining Techniques (Decision Tree Models) for Examining the Impact of Internet-Based Cognitive Behavioral Therapy for Tinnitus: Machine Learning Approach
J Med Internet Res 2021;23(11):e28999
URL: https://www.jmir.org/2021/11/e28999
doi:10.2196/28999
PMID:34726612

Original Paper

Adaptation of a Theory-Based Social Networking and Gamified App-Based Intervention to Improve Pre-Exposure Prophylaxis Adherence Among Young Men Who Have Sex With Men in Bangkok, Thailand: Qualitative Study

Wipaporn Natalie Songtaweesin[1], MBBS, DTMH[‡]; Sara LeGrand[2], PhD; Shashika Bandara[2], MSc; Caitlin Piccone[2], MPH; Prissana Wongharn[1], MSc; Juthamanee Moonwong[1], MSc; Thidarat Jupimai[1], MSc; Chutima Saisaengjan[1], MA; Tuangtip Theerawit[1], MPH; Kathryn Muessig[3], PhD; Lisa Hightow-Weidman[3,4], MPH, MD; Thanyawee Puthanakit[1,5], MD, MHS; Nittaya Phanuphak[6], MD, PhD; Arunrat Tangmunkongvorakul[7], PhD

[1]Faculty of Medicine, Center of Excellence for Pediatric Infectious Diseases and Vaccines, Chulalongkorn University, Bangkok, Thailand

[2]Center for Health Policy and Inequalities Research, Duke Global Health Institute, Duke University, Durham, NC, United States

[3]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[4]Behavior and Technology Lab, Institute for Global Health and Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[6]Institute for HIV Research and Innovation, Bangkok, Thailand

[7]Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand

[‡]The CE-PID - IHRI Adolescent Study Team

Corresponding Author:
Arunrat Tangmunkongvorakul, PhD
Research Institute for Health Sciences
Chiang Mai University
110 Intavarorot Road, Tambol Siphum
Amphur Muang
Chiang Mai, 50200
Thailand
Phone: 66 815942773
Fax: 66 53221849
Email: arunrat@rihes.org

Abstract

Background: HIV disproportionately affects young Thai men who have sex with men (YMSM). Recent studies report a high incidence and prevalence of HIV among Thai YMSM. The Thai national guidelines have recommended pre-exposure prophylaxis (PrEP) since 2014 for key populations; free PrEP has been piloted since 2019. Smartphone-based mobile health (mHealth) interventions provide an optimal platform for innovative PrEP adherence interventions for Thai YMSM.

Objective: This study aims to adapt the P3 (Prepared, Protected, emPowered) app, developed with YMSM and transwomen in the United States to improve PrEP adherence and persistence for YMSM in Thailand. The app aims to provide daily adherence support and addresses gaps in staff available for large-scale PrEP rollout needed to see population-level effects of HIV prevention.

Methods: We conducted focus group discussions (FGDs) with YMSM and key informant interviews (KIIs) with PrEP care providers in Bangkok, Thailand, to investigate PrEP adherence facilitators and barriers, preferences for functions and features in mHealth apps among YMSM, and how to best adapt the P3 app to the Thai context. We conducted four FGDs with 4-8 participants per group and 15 KIIs.

Results: For FGDs, 23 YMSM participated with a mean age of 20 years (range 18-21), 96% (22/23) enrolled in full-time education, and all owned smartphones. The mean age of KII participants was 40 (range 26-60) years; most were state health service providers, with the majority being counselors (6/15, 40%) and physicians (6/15, 40%). Overall, the facilitators and barriers for PrEP adherence identified were similar to those of MSM and YMSM globally including the United States. Key themes included

general recommendations for improving mHealth apps in Thailand, such as presenting reliable information in an appealing format, minimizing privacy risks, and addressing connectivity challenges. Additional themes focused on P3 Thailand adaptations and were related to cultural and stylistic preferences, engagement strategies, and recommendations for new functions. To develop the adapted app, P3 Thailand, these findings were balanced with resource limitations resulting in the prioritization of minor modifications: changes in app esthetics (color scheme, iconography, and imagery) and changes in the presentation of information in two of the app's features. FGDs identified similar PrEP adherence facilitators and barriers to those already addressed within the app.

Conclusions: The core elements of the P3 app address major PrEP facilitators and barriers for Thai YMSM; however, changes to the app features, including stylistic presentation, were needed to appropriately customize the app to the Thai context. Given the similarities of facilitators and barriers for PrEP adherence globally, adapting existing PrEP mHealth solutions based on input from end users and key informants provides a promising approach. However, partnerships with local app designers and developers can improve the adaptation process and final product.

Trial Registration: ClinicalTrials.gov NCT04413708; http://clinicaltrials.gov/ct2/show/NCT04413708

(J Med Internet Res 2021;23(11):e23852) doi:10.2196/23852

KEYWORDS

mobile health; young men who have sex with men; pre-exposure prophylaxis; adherence; mobile phone

Introduction

In Thailand, gay, bisexual, and other men who have sex with men (MSM) are disproportionately affected by HIV [1]. Studies conducted in Bangkok have identified a high HIV incidence and prevalence among MSM, with young MSM (YMSM) at a significantly greater risk of HIV infection than older MSM [2-4]. Examinations of data from YMSM aged 18-24 years enrolled in the Bangkok MSM Cohort Study (40.8%; n=712) between 2006 and 2014 found a baseline HIV prevalence of 21.2% and an overall HIV incidence of 7.4 per 100 person-years, a rate much higher than that for MSM of all ages (5.3 per 100 person-years) [3,4]. To address disparities in HIV infection among YMSM in Thailand, the implementation of novel behavioral and biomedical interventions that are culturally and developmentally appropriate, useful, and scalable are urgently needed [4,5].

Pre-exposure prophylaxis (PrEP), a combination drug containing tenofovir and emtricitabine taken once daily, is a safe and effective method for preventing transmission of HIV among MSM [6-8]. Owing to the high levels of efficacy found in PrEP trials, the Thai National Guidelines on HIV/AIDS Treatment and Prevention in 2014 were revised to recommend PrEP for key populations, including YMSM [9,10]. Since October 2019, a coverage scheme providing free PrEP for key populations, including YMSM, has been piloted in Thailand's National Health Security Office, with a view for future scale up [11,12]. Although PrEP availability is expanding rapidly for Thai YMSM, PrEP effectiveness is highly correlated with adherence, and evidence suggests poorer adherence among YMSM, including those in Thailand [7,13-16]. For example, two studies evaluating the Princess PrEP program, the first key population-led PrEP initiative in Thailand, found lower levels of PrEP adherence among MSM participants aged <25 years compared with those aged >25 years [13,14].

Smartphones provide an optimal platform for the delivery of an innovative PrEP adherence intervention in addition to other chronic medical conditions in adolescents because they are an integral part of their lives, including Thai YMSM [14,17-21]. There is growing evidence to support the improvement in self-management related outcomes in adolescents using electronic health interventions, in addition to user feasibility, acceptability, and satisfaction of such interventions [20,22,23]. Data collected by the Thai Red Cross AIDS Research Center (TRCARC) from 2015 to 2017 on the use of social networking sites by MSM aged >18 years and at risk of HIV (n=465) showed that users were satisfied with web-based service delivery primarily delivered through smartphones [24]. In addition, app-based HIV prevention interventions have shown high rates of acceptance among MSM, including YMSM, in high-income countries and have potential use in low-middle income countries [25-29]. On the basis of high smartphone use and findings from other study teams in Thailand and globally, app interventions are likely to be acceptable, help to overcome barriers to engagement with in-person interventions such as inconvenience, transportation, and stigma; and address challenges associated with medication, including PrEP adherence and other HIV prevention behaviors among adolescents, including YMSM in Thailand [13,14,17,22-24,30].

Rather than building de novo smartphone HIV prevention apps, there is growing support for adapting and building on existing evidenced-based platforms that share common behavior change goals, theoretical underpinnings, and features to maximize the potential for intervention sustainability, affordability, and scalability [31-33]. As such, this study adapted the theory-based P3 (Prepared, Protected, emPowered) app designed by researchers at the University of North Carolina at Chapel Hill and Ayogo Health, Inc [33]. The P3 app aims to improve PrEP adherence, retention in PrEP clinical care, and PrEP persistence among YMSM and young transgender women who have sex with men in the United States. To achieve its aims, P3 uses social networking and game-based elements along with numerous evidenced-based features, including social cognitive theory, narrative communication, and principles of the Fogg behavioral model of persuasive technology to promote behavioral change through triggers, ability, and motivation [34]. The purpose of this study is to conduct formative research to

guide adaptations to P3 that will optimize app relevance, user engagement, and utility of the adapted app for Thai YMSM, henceforth known as P3 Thailand (P3T). We hypothesize that stylistic and cultural preferences will be different from US YMSM preferences, particularly with app esthetics and preference for more visual content and less text.

Methods

Overview

To adapt the P3 app to the Thai context, we collected qualitative data through focus group discussions (FGDs) with YMSM and key informant interviews (KIIs) in Bangkok, Thailand, between January and March 2019. Participants in both the FGDs and KIIs (1) described facilitators and barriers to PrEP adherence among Thai YMSM, (2) discussed mobile health (mHealth) app features and functions preferred by YMSM; and (3) provided specific feedback about P3 and offered suggestions for adapting it to create P3T.

Institutional review board approval was granted for this study by the Faculty of Medicine, Chulalongkorn University, with a waiver for parental consent granted. This study was registered with ClinicalTrials.gov (NCT04413708).

All study participants completed a brief computer-assisted self-interviewing survey on patient demographics, technology use, and previous experience using or providing PrEP before their session. All FGDs and KIIs were facilitated by a study staff member using a semistructured interview guide. After exploring PrEP adherence facilitators and barriers and mHealth app preferences among Thai YMSM, the facilitator gave a presentation of key features of P3 and elicited feedback and suggested changes from study participants.

The presentation of P3 features included screenshots and descriptions translated into Thai by the facilitator or interviewer starting with screenshots of onboarding activities (eg, review of community guidelines, app terms and conditions, setting up an account, option to lock screen with a PIN, brief introductory screens describing key P3 features, and profile setup). Next, participants were shown a screenshot of the user dashboard, or home screen, with an explanation of each icon. Five key features of P3 (Textbox 1) were demonstrated using screenshots and explained in detail by the facilitator, including the social wall, daily quests, medication tracking and adherence support, knowledge center, and collections. An additional key feature, adherence counseling, was not presented via screenshots but was explained during the presentation of the user dashboard screenshot. A brief description of these key features is provided in Textbox 1. Detailed information on all P3 features has been described elsewhere [34].

Textbox 1. P3 features discussed in focus group discussions and key informant interviews to adapt for Prepared, Protected, emPowered Thailand.

Key Prepared, Protected, emPowered (P3) features discussed in detail in focus group discussions and key informant interviews:

Social wall:

- Daily prompts (eg, "Share tips for being safe if you are having a lot of sex!") provide a safe space for discussions on topics related to pre-exposure prophylaxis (PrEP) adherence and sexual health such as mental health and positive sexual health habits
- Peer sharing, support, and reinforcement are intended to foster behavior change and build a sense of community among users
- Users are sent push notifications when someone has commented on or *liked* a post

Daily quests:

- Routine tasks that help users change behaviors by building knowledge and skills and setting goals (eg, "Why did you decide to start taking PrEP? Consider your reasons for starting PrEP and write them down. Knowing your reasons for taking PrEP can help motivate you!")

Medication tracking and adherence support:

- To promote medication-taking habits, users personalize adherence strategies based on the time or circumstances in which they plan to take their medication (eg, "I take PrEP when I brush my teeth")
- The medication (med) reminder system provides discreet and focused reminders based on user-selected time and adherence strategies. Reminders are linked to a med check prompt that allows users to track adherence or nonadherence for the past 3 days; data entered are used to update the in-app calendar, which shows adherence history. Users can also access the med check prompts and adherence history feature from the home screen
- Adherence support: The study team can monitor users' PrEP adherence and app use. Encouraging messages can be sent to the user to provide support and encouragement

Knowledge center:

- A multimedia library with information on PrEP use, frequently asked questions about PrEP, PrEP side effects, and sexual health–related topics. A progression bar shows completion of each section
- Users can read articles to build their knowledge, respond with one of three responses indicating the usefulness of the article, and answer questions that encourage personal reflections about the content (not visible to other app users)

Collections:

- A choose-your-own-adventure style story for the user that closely mirrors choices and circumstances they may face in their daily lives
- Playing through these storylines allows users to build problem-solving skills when faced with difficult choices
- The *collections* feature also acts as an incentive for users to engage with other features of the app to gain points to unlock stories

Adherence counseling:

- Personalized adherence counseling is provided via secure in-app messaging; addresses adherence challenges faced by young men who have sex with men
- Key features include reviewing sexual risk experiences of the participants, exploring PrEP adherence facilitators and barriers, identifying adherence needs and strategies to meet needs, discussing mental health and relationship issues, and developing an adherence action plan [35,36]

Recruitment for FGDs

FGD participants were recruited from (1) HIV prevention clinics at the Center for Excellence for Pediatric Infectious Diseases at the Chulalongkorn University, a major teaching hospital in Bangkok; (2) TRCARC, Thailand's largest HIV voluntary counseling and testing (VCT) center, which serves over 2000 adolescents per year; and (3) Two community-based organization branches of TRCARC, Service Workers IN Group and Rainbow Sky Association Thailand. Potential participants were approached at HIV VCT visits to assess their interest in study participation. In addition, study advertisements were posted in clinics, social media, and via the TRCARC website. The inclusion criteria were as follows: (1) age 16-24 years, (2) male sex at birth, (3) self-identification as MSM, (4) ability to speak and read Thai, (5) familiarity with smartphones, and (6) current or previous PrEP use. Those unable to provide consent due to substance use or psychological conditions were excluded.

FGDs lasting 60-120 minutes were conducted in private conference rooms either at offices in clinics or reserved private conference rooms in cafes in groups of 4-8. All participants were identified using a pseudonym to protect their confidentiality. At the conclusion of the FGDs, participants were provided THB 500 (US $16) to compensate them for their time.

Recruitment for KIIs

Key informants who had collaborated with the Center for Excellence for Pediatric Infectious Diseases or participated in academic conferences on adolescent health in Bangkok were recruited for KIIs via in-person contact, phone calls, text messages, and email invitations. To participate in a KII,

individuals were required to be a current PrEP provider for YMSM in Bangkok. KIIs, which lasted between 40 and 60 minutes were conducted face-to-face at PrEP clinics. The KII participants were not compensated.

Data Analysis

Pre-FGD and KII survey data were summarized using Adobe Acrobat. FGDs and KIIs were audio-recorded and transcribed in Thai. A directed content analysis approach, defined as categorization derived directly from text data and transcribed interviews, was used [37]. An initial codebook was developed collaboratively between the US and Thai study teams based on semistructured interview guides that reflected the study research themes and subthemes. An iterative approach was used when new codes arose, as interviews were conducted. New codes emerging were added, and any transcribed code that had already been coded was recoded. Two members of the Thai study team coded the FGD and KII transcripts using NVivo (version 12.5.0) [38]; statistical interrater reliability for coding was not calculated. Instead, divergence among coders was discussed among the analysis groups to reach consensus, resolved by the Thai site principal investigator. Data were then deductively analyzed using thematic analysis by the same Thai study team members to identify key themes and subthemes relevant to the adaptation of P3 for YMSM in Thailand. Quotes illustrative of key themes were identified and translated into English by bilingual Thai English speakers.

Results

Participant Characteristics

Characteristics of FGDs

A total of 23 YMSM participated in four FGDs. Participants were aged 18-21 years, with a mean age of 20 years. The majority were enrolled in full-time education 96% (22/23), with 70% (16/23) enrolled in university. All participants owned a smartphone and used it to access the internet; 74% (17/23) were on self-paid monthly internet plans. Over half of the participants (13/23, 57%) had phones that used an iOS (Apple iPhone operating system) and 35% (8/23) used an Android operating system. Most (22/23, 96%) were current PrEP users and 4% (1/23) were previous PrEP users.

Characteristics of KIIs

A total of 15 KIIs were conducted; participants were aged 26-60 years with a mean age of 40 years. Most were counselors (6/15, 40%) and physicians (6/15, 40%); the remaining 20% included 2 nurses and 1 social worker. All patients were employed at publicly accessible HIV VCT clinics. The majority of key informant interviewees had completed university level education (11/15, 73%), and 60% (9/15) had been providing PrEP to adolescents for 3-4 years.

Overview of FGD and KII Findings

General Overview

Key findings from the FGDs and KIIs were grouped into two main categories. The first focuses on the facilitators and barriers to PrEP adherence among YMSM in Bangkok. The second focuses on general and specific recommendations for ensuring P3T app relevance, user engagement, and utility for Thai YMSM. We present these themes and subthemes along with English translations of illustrative excerpted participant quotes.

PrEP Adherence Facilitators and Barriers

FGD Findings: Adherence Facilitators

Key subthemes that emerged regarding facilitators of PrEP adherence discussed by participants included risk perception, PrEP cost, and social support.

Risk Perception

FGD participants who felt they were at risk of HIV infection were more motivated to take PrEP:

> *I had a pretty risky history before. I use my history to determine my risk so now take it (regularly). I also take it because there were people that recommended it to me.* [FGD1, Participant 3]

PrEP Cost

PrEP being available free of charge was also important in supporting PrEP adherence:

> *It really helps with us being students [that PrEP is free] as it helps us save...if we had to pay for it, I would have to think a lot more whether or not to take it.* [FGD4, Participant 6]

Social Support

Social support from staff was another subtheme that FGD participants gave importance to in their willingness to adhere to PrEP:

> *For me what's important is staff that are constantly available for counseling. If we had providers that just gave us our meds and that was the end of it—no questions about how we were, I wouldn't want to come for PrEP services.* [FGD4, Participant 3]

KII Findings: Adherence Facilitators

The key facilitators of PrEP adherence discussed by key informants were similar and included social support as illustrated in the quote below.

> *For some couples who both take PrEP and are open to each other about this and have an open relationship, taking PrEP in front of each other does not need to be hidden for fear of being accused of being unfaithful.* [KII, Participant 14]

FGD Findings: Adherence Barriers

Subthemes that arose on the issue of major barriers for YMSM in PrEP adherence included stigma, logistical issues, and PrEP side effects.

Stigma

YMSM felt that taking PrEP was associated with stigma, risky sexual behavior, being gay, and also being seen to be infected with HIV:

> *People will feel that [PrEP] clinics are for gay men with risky behaviors. It's well known in Bangkok that a lot of gay men are infected with HIV...people look at gay men negatively as it is...that we're promiscuous.* [FGD4, Participant 5]

Logistical Issues

Logistical issues, including being busy with other activities, also featured prominently in reasons YMSM felt it was difficult to always adhere to PrEP:

> *We have a lot of classes and have to study for exams, there isn't always time to come for appointments.* [FGD2, Participant 3]

PrEP Side Effects

PrEP side effects were another issue adolescents felt made PrEP adherence challenging, particularly when it interfered with other activities in their daily lives:

> *I had a lot of side effects with PrEP, and I have to study everyday—it makes it impossible to study.* [FGD3, Participant 1]

KII Findings: Adherence Barriers

Findings of major barriers from the point of view of KIIs were very similar to those of YMSM as already discussed above, including logistical issues, risk perception, and peer influence.

KIIs experienced difficulties arranging logistics to come for appointments and low HIV risk perception, in addition to a number who initiate PrEP because of peer influence and discontinue it because of lack of intrinsic motivation or sense of self-risk perception for HIV:

> *Many kids who don't turn up to appointments will say its inconvenient for them to come in, they don't have time to come, or they are no longer at risk. Some say they want to come but borrow PrEP from friends...confess they don't have time for appointments but initially came because they saw their friends taking PrEP and wanted to join in, we see this a lot.* [KII, Participant 2]

Adaptations Made Based on PrEP Adherence Facilitators and Barriers

As the facilitators and barriers found in this study were essentially the same as those seen in the US P3 app, no changes were made to the app regarding this [34].

General and Specific Recommendations for Ensuring P3T App Relevance, User Engagement, and Utility for Thai YMSM

Six main themes emerged under this category: (1) information provided in mHealth apps, (2) stylistic preferences, (3) engagement strategies, (4) privacy risks, (5) connectivity challenges, and (6) recommendations for new functions. The first 3 themes focused on general recommendations for mHealth apps in Thailand based on discussions before the review of P3 features. The next 3 themes specifically focused on feedback and suggested adaptations to make the app more relevant to Thai users based on the review of P3 features.

Information Provided in the App

Three main subthemes related to app information emerged: (1) access to reliable information, (2) presentation of information, and (3) space for peer influence and information sharing.

Access to Reliable Information

In FGD findings, Several FGD participants discussed the ways in which Thai societal norms can be restricted to YMSM seeking information related to sexual health. However, they felt that access to such information was vital to self-care:

> *I think an app will help us feel less afraid to ask...sometimes society leads [us] to believe certain things are not okay [to ask] but it is actually beneficial. We want more freedom to be able to get hold of simple information to be able to look after ourselves. A lot of us worry looking after ourselves is too difficult and complicated.* [FGD1, Participant 5]

Despite the vast amount of information about PrEP available in Thai on the internet and social media sites, FGD participants were concerned that many sources lack reliability:

> *...it's better to have accurate information, for example, answers to questions from doctors, [but] it's so difficult to find official medical websites. Most of what I find is opinion sharing of the general public [on social media]...this is a problem.* [FGD3, Participant 2]

Presentation of Information

FGD Findings

Most FGD participants commented on the importance of the presentation of information in increasing app interest and engagement. They indicated a strong preference for information that was presented concisely and was visually oriented:

> *I think [information] should be [presented as] small nuggets of information, and when you click on it, it takes you to...an infographic, like a poster with concise information.* [FGD1, Participant 1]

> *Most teenagers will stop to read if there are graphics, but they will not read it if there is just text.* [FGD2, Participant 6]

In addition, participants highly recommended including video content and links to useful content outside the app.

> *When I open anything in Facebook, I see some interesting clips on pages for gay people—they will provide information about this stuff. If they have a collection of links of places to chill out, or links to You Tube videos, it just makes it more interesting.* [FGD1, Participant 2]

KII Findings

Another presentation style suggested was the use of social media influencers who use diverse methods to communicate information that is more appealing to YMSM:

> *I think having influencers on there with different messages. I've seen pictures on Instagram and Facebook and I think going there and being able to see their favorite movie star or whoever influences the community.* [KII, Participant 11]

Peer Influence and Information Sharing

FGD Findings

Participants noted the strong influence peers have on their decision-making; therefore, peers were viewed as key facilitators for app uptake and sustained engagement:

> *Thai teenagers listen much more to the opinion of their friends than making decisions on their own.* [FGD3, Participant 4]

> *I think if teens can see what other people think [is] worth reading, it would make me like, oh that looks like something I want to read, why are there so many people reading this?...like getting recommendations from others about it.* [FGD 1, Participant 4]

> *Sometimes teenagers may just want to talk to their friends or want information from them,-for example, who is taking PrEP and how they found it. People with previous experience will flood in to answer these queries, like on how they manage side effects. They tend to want experiential information from real-life users, which they find more convincing than information from doctors.* [KII, Participant 2]

KII Findings

Key informants also discussed the powerful nature of platforms for youth to connect with one another. They have the potential to provide a safe space for youth to discuss and share information about health-related topics that may be difficult to talk about in other settings while also creating opportunities to discuss similar life experiences with others:

> *I once had a closed online group for teenagers—this group were constantly talking. They would talk about health, then they would talk about something else, then back to health and medicines...an app could be potentially great for teens supporting each other towards a shared goal.* [KII, Participant 1]

Stylistic Preferences

FGD Findings

The presentation of key P3 features in the FGDs led to rich discussions about the *look and feel* of the app and its features. The app screenshots presented to study participants showed the P3 superhero theme, which included sleek, abstract superhero avatars, and a darker background color scheme. These features were found to be highly acceptable in formative research conducted in the United States [34,39]. FGD participants nearly unanimously agreed that the theme and design elements were not aligned with app preferences among Thai youth.

> *I think the color scheme is a little too dark, it should be more colorful and cute...there should be a wider choice of color schemes.* [FGD1, Participant 7]

> *The avatars look really mysterious...strange...I want the avatars to look more like us, like ones that can wave...cartoons...that we can choose...or use emojis that look more normal.* [FGD1, Participant 5]

YMSM also expressed the need for cartoon-style stickers, similar to other mainstream platforms used in Thailand, such as LINE, the most frequently used instant messaging service in Thailand [39]. They noted that they are a prominent part of contemporary communication in Thai culture; hence, they should be featured in a PrEP adherence app.

Furthermore, participants indicated a preference for the social wall feature of P3T to have a more attractive look and feel, similar to current popular social networking sites such as YelloTalk (Figure 1):

> *I want the Social Wall to look like YelloTalk.* [FGD 3, Participant 4]

Figure 1. Yellotalk screenshots as displayed in the Apple App Store.

KII Findings

KII participants commented in a similar way to FGD participants, expressing that youth would likely enjoy a lighter color scheme and a cuter style of graphics:

> *This looks too depressing and dark...a lighter color theme would be better...the avatars need to be cute cartoons for the kids to pick...* [KII, Participant 3]

KII participants noted that electronic stickers are a prominent part of contemporary communication in Thai culture; hence, they should be featured in a PrEP adherence app:

> *Here, everybody is about LINE and stickers. I even engage with senior ministry staff by line stickers. So, it's not only kids, it's everybody.* [Participant 11]

Engagement Strategies

FGD Findings

FGD participants provided several important recommendations for improving their engagement with the P3T app. First, the app content should be highly dynamic; suggestions include regularly adding new articles, posts about upcoming social events, tailoring to special dates, and current news tailored to individual users and adolescents. Notifications should be sent to alert users to the availability of new content, including contributions from other app users:

> *If there is an event somewhere it should notify us or notify us about interesting news that will make us want to click in to see. If there is nothing going on in the app and it's just a boring app, I have several apps on my phone, I would just scroll past.* [FGD1, Participant 2]

> *I look at what updates there are, whether they are interesting. There needs to be exchange of knowledge and notifications to make it more interesting, like notifying us a new post has been made on something that interests us, tailoring it to our interests to make us want to use it more.* [FGD3, Participant 4]

The importance of fun was emphasized by FGD participants, as they highlighted the need for more entertaining features than those found in the reviewed P3 features to optimize engagement:

> *Something that has a changeable theme, has fun gimmicks, like stress busting games...like the games you can get in Facebook will make the app more appealing.* [FGD1, Participant 8]

Study participants had varying opinions about rewards that may best incentivize user engagement with the app. Several felt that Thai YMSM would be most interested in earning points for completing app activities that could be redeemed for external rewards such as cash, tickets, or discounts for services.

> *Apps that I usually use have daily questionnaires...they give point rewards...we get gift vouchers...[we like to exchange points for]...bubble tea, cash, movie tickets, discounts for spas and botox, fastfood, cafes, shoes, buying games.* [FGD2, Participant 1]

Some FGD participants expressed interest in the choose-your-own-adventure style storytelling collections feature and suggested increasing the adventure and excitement of the storylines. However, to be effective as an incentive for app use, storylines would need to be more exciting and adventurous.

> *We enjoy reading stories about unique experiences of others...oral sex, group sex...* [FGD 1, Participant 2]

KII Findings

When asked about factors that should be taken into account to maximize engagement in a PrEP adherence app for Thai YMSM, a key informant provided critical insight into the importance of fun in Thai culture:

> *There is a strong need in the Thai context of interaction for fun. It has to all be about carrots not sticks. About fun. The Thai culture is about—what's the point if you're not going to have fun along the way.* [KII, Participant 11]

A key informant commented specifically about users being limited to redeeming their earned points in P3T to buy choose-your-own-adventure stories in the collections feature. The participant was not optimistic that this approach would incentivize app engagement among Thai YMSM:

> *Our teenagers don't have a lot of persistence to do things. If they have to go in to read something, it won't really motivate them to want to know more, unless it comes in the form of a game, and they get points for doing it...point exchange needs to be for something they want...using points to buy something to read [referring to Collections], I don't know, that's not really in the nature of Thai teenagers to do something like that.* [KII, Participant 9]

Privacy Risks

Many FGD participants expressed concerns about sharing their private information on the web and emphasized their preference for personal interactions with health care providers:

> *I would rather talk to doctors and counselors than tell my personal things to friends that I don't know.* [FGD4, Participant 6]

On the basis of the fear of private information being compromised, some FGD participants expressed the need for advanced methods of user authentication:

> *I want a highly secure system, maybe use a finger or face scan...to access the app...because our medication tracking information and blood results are sensitive information.* [FGD1, Participant 4]

Connectivity Challenges

A main challenge specific to the use of mHealth apps in Thailand is reliable internet access. Many apps, including P3T, require a continuous active internet connection to function. However, internet connectivity is not readily available or affordable for many Thai adolescents. The FGD participants noted the importance of offline app functionality:

> *For me, like if I don't have internet this week, or I haven't bought internet time or haven't gone home this week, I can fill in my questionnaire responses...and when I'm back online, my survey responses are sent.* [FGD 4, Participant 8]

Recommendations for New App Functions

FGD Findings

In FGDs, two additional app functions were discussed most frequently and recommended for inclusion in the P3T. The first recommendation was to add a search function that allowed users to search all of the app's content:

> *When I do an internet search [for information] there are so many websites, and sometimes they are focused on selling you things...it takes ages to find the information I actually want...information in an app will get this to information I want directly.* [FGD3, Participant 4]

The second recommendation was to include a function that allows users to monitor and plan a healthy lifestyle. For example, FGD participants wanted the ability to track HIV risk behaviors in the app and receive customized information about their HIV risk level based on their tracked data:

> *I want it [the app] to help determine [my risk] based on data I have inputted...if I had a risk event today...how should I deal with it...and use it to go through events with my doctor when I next visit them.* [FGD1, Participant 5]

KII Findings

Key informants pointed out that Thai YMSM appreciate health-related feedback from adolescent-friendly providers, and offering individually tailored information through P3T could be valuable:

> *One of the things we hear back from some of our participants in research is that they like having health information. We noticed that they're interested in coming back to their visits so they can get health information and feedback on how they're doing health wise. I think that can potentially be interesting to people. "Your kidney health is great" and "you don't have any evidence of STI that's awesome" or "keep up the good work" or whatever it is.* [KII, Participant 11]

Adaptations Made for P3T Based on Formative Research Findings

On the basis of formative research findings and the limited resources available for adaptations, we prioritized changes to app features while keeping the same content to address the presentation of information provided in P3T and some identified stylistic preferences. These included (1) changing the information presentation style in the knowledge center, (2) adapting stories and images in the collections feature, and (3) changing avatar options. Longer-term structural changes that were not immediately feasible because of resource constraints should be considered in future updates to the app.

Presentation Style Changes in the Knowledge Center

On the basis of feedback, we changed our information presentation style by creating chapter images more relevant to the Thai context (Figure 2), changing the text-based presentation of information to include less text and more infographics (Figure 3), and added links to outside video content, including those featuring social influencers (Figure 4). All of these videos were locally produced by the Thai HIV prevention NGO, Love Foundation, and reflected popular Thai presentation styles.

Figure 2. Comparison of knowledge center chapter images before (left) and after (right) adaptation for the Thai context.

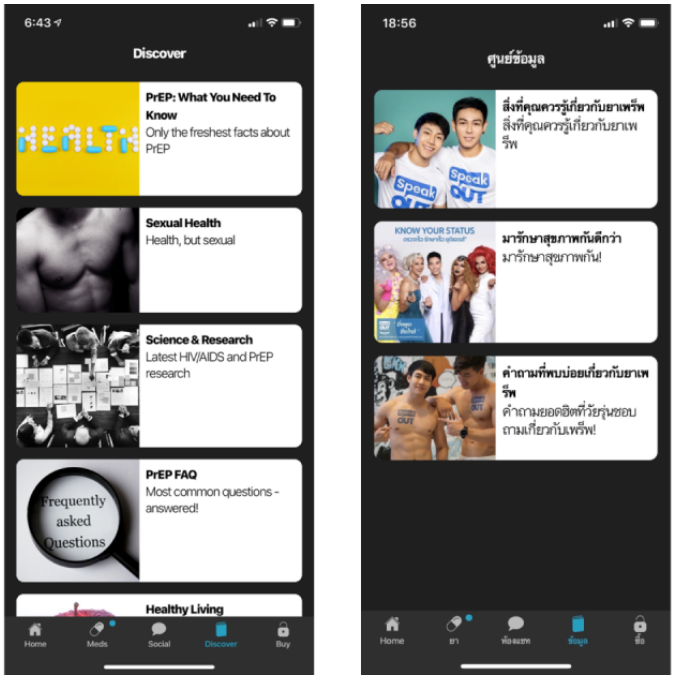


Figure 3. Comparison of knowledge center article on pre-exposure prophylaxis adherence before (left) and after (right) adaptation for the Thai context, including infographics.

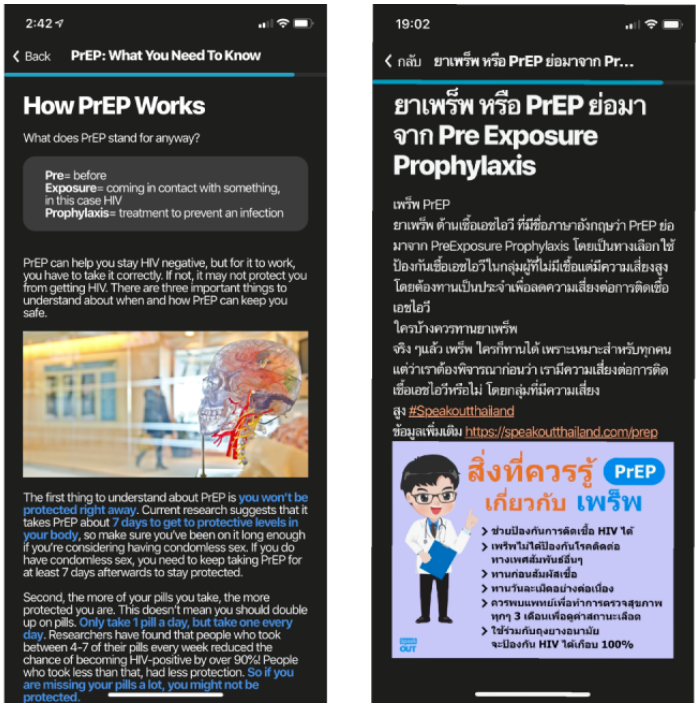
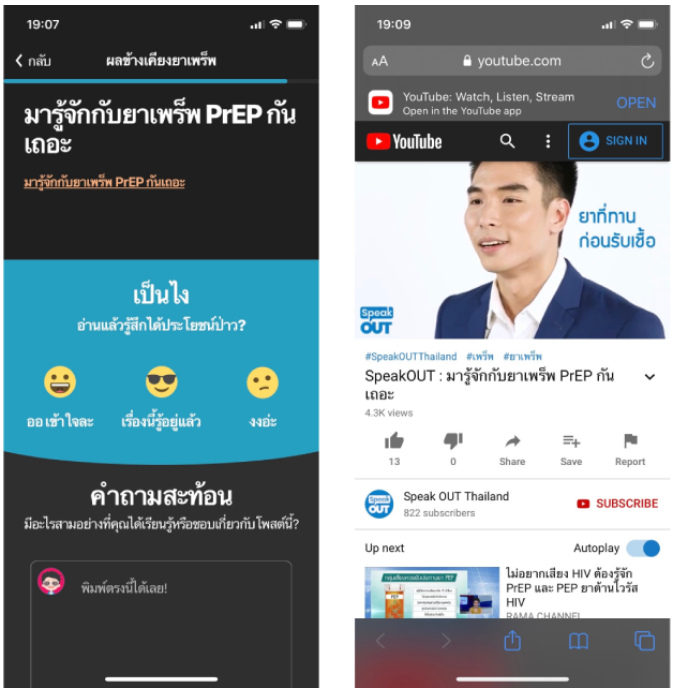


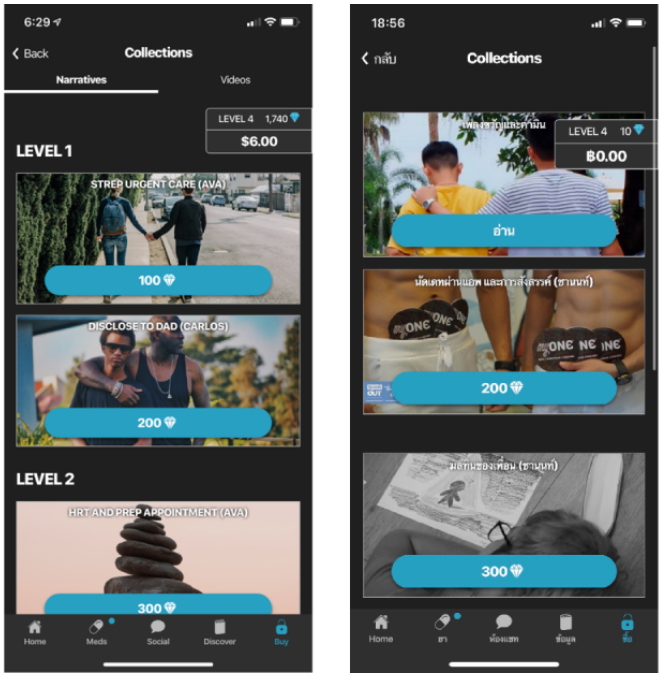
Figure 4. Postadaptation knowledge center page (left) with link to external video content (right).



Adaptation of Stories and Images in the Collections Feature

The collections feature was adapted by giving story characters Thai names and locations and changing some story lines to increase relevance for Thai YMSM. In addition, image icons for the collection titles were changed to Thai characters (Figure 5).

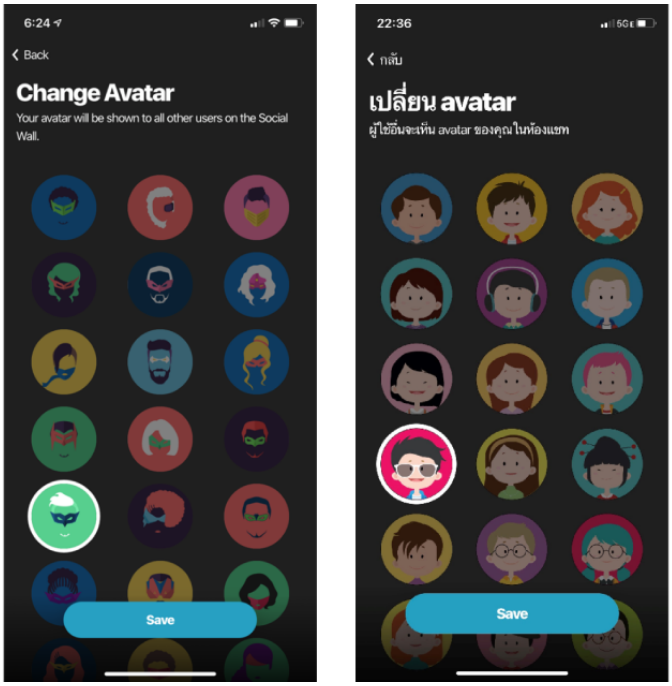
Figure 5. Comparison of Collection titles before (left) and after (right) adaptation for the Thai context.



Avatar Adaptations

As indicated by many FGD participants, the original avatars presented did not appeal to Thai YMSM and were considered a major detraction of the app. We partnered with Thai graphic designers to design a new set of avatars to replace the original avatars. Figure 6 compares the older version and the new version of avatars.

Figure 6. Comparison of avatars before (left) and after (right) adaptation to the Thai context.



Existing P3 Features That Foster PrEP Adherence Facilitators, Address Adherence Barriers, and Align With Key Recommendations

In addition to the specific changes made as indicated in the previous section, many P3 features already address PrEP adherence facilitators and barriers and general app recommendations identified in FGDs and KIIs. For example, P3 seeks to increase PrEP adherence facilitators by providing content that promotes accurate self-assessment of HIV risk via knowledge center content, promoting the exchange of support and information on topics related to PrEP adherence via the social wall, and offer personalized adherence support from a trained adolescent-friendly adherence counselor through the in-app counseling feature.

The knowledge center and social wall also already address some of the general app recommendations related to the reliability of

information, access to information, and peer information sharing. The knowledge center is designed to provide reliable information in an easy-to-understand manner. The social wall provides a space for users to discuss topics related to health anonymously. In addition, in terms of security and privacy concerns, the P3 app provides an in-app log-in option for the user to secure the app either by using a personalized PIN number for all platforms or using face recognition technology for iOS platform users when launched. The app helps to protect personal identity by asking users to set their own pseudonyms that do not contain any identifying information about them rather than using their real names in the app. A summary of key study findings and adaptations made is made in Table 1.

Table 1. Summary of key study findings and adaptations made.

Findings	Relevant quotes	Adaptations made
Presentation style preferences		
• Images more reflective of Thai YMSM[a] settings • Preference for more visually based information and less text • Preference for information from social influencers • Preference for a cuter style and lighter color theme • Content needs to reflect their local cultural context	• "I want the avatars to look more like us." [FGD[b]1, P[c]7] • "I think [information] should be [presented as] small nuggets of information...an infoographic, like a poster with concise infographic." [FGD1, P1] • "I think having influencers on there with different messages." [KII[d], P11] • "A lighter color would be better...the avatars need to be cute cartoons." [KII, P3] • "I want the Social Wall to look like YelloTalk." [FGD3, P4]	Knowledge center • Graphics incorporated more photos taken in the Thai YMSM context • More infographics and less text used • Links added to outside video content containing messages from social influencers Collections • Image icons changed to Thai characters • Story characters given Thai names and locations Avatars • Created by Thai graphics designer with cute theme and lighter color scheme
Facilitators		
• High HIV risk perception, PrEP cost support, social support	• "I used my [risky] history to determine my risk." [FGD1, P3] • "It really helps with us being students [that PrEP is free." [FGD4, P6] • "Some couples who both take PrEP are open to teach other." [KII, P14]	• No adaptations made as same facilitators and barriers as found in P3 US study results
Barriers		
• Stigma, logistical issues, PrEP side effects, low risk perception, and peer influence	• "People look at gay men negatively as it is." [FGD4, P5] • "We have a lot of classes...there isn't always time to come for appointments." [FGD2, P3] • "I had a lot of side effects...makes it impossible to study." [FGD3, P1] • "Some came because they saw their friends taking PrEP...but confess they don't have time for appointments." [KII, P2]	• No adaptations made to current version because of budgetary restrictions
New app function recommendations		
• Search function and risk behavior tracker	• "Information in an app to get information I want directly." [FGD3, P4] • "I want it to help determine [my risk]." [FGD1, P5]	• No adaptations made to current version because of budgetary restrictions

[a]YMSM: young men who have sex with men.

[b]FGD: focus group discussions.

[c]P: participant.

[d]KII: key informant interview.

Discussion

Overview

This study explored facilitators and barriers related to PrEP adherence and elicited general and specific recommendations from Thai YMSM and key informants to ensure that adaptations of a PrEP adherence app originally developed for YMSM in the United States were relevant, engaging, and useful for the Thai context. To our knowledge, this is one of the few studies conducted outside high-income settings that address the adaptation of an mHealth approach to support PrEP adherence among YMSM [40].

The individual, societal, and systemic level facilitators and barriers to PrEP adherence among Thai YMSM identified in this study have also been reported in other studies of MSM and YMSM in the United States, Thailand, and other countries. Consistent with other studies [7,41-45], adherence facilitators included awareness of HIV risk and desire to protect themselves, concerns related to sexual partners, support from friends, sexual partners, family members, and adolescent-friendly providers, and PrEP affordability (PrEP is provided free of charge through universal health coverage for key populations in Thailand). Key barriers, also identified in other studies [41-44,46-49], included sexuality, HIV and PrEP stigma, attending regular PrEP medical appointments, forgetting to take PrEP daily or while intoxicated, not having PrEP with them at certain times, running out of PrEP, side effects, and low HIV risk perception. Because the PrEP facilitators and barriers identified in this study had been identified in prior research and were already addressed by the original P3 app, its content was already designed to foster these facilitators and address the barriers and consequently no content was changed; only app feature changes were made in adaptations in this trial.

In general, the main features, such as the social wall, daily quests, and collections of the P3 app were found to be acceptable among the study participants. It is likely that this is in part because of the ways in which the key app features address the factors identified by participants in this study as the most important PrEP adherence facilitators and barriers among YMSM in Thailand. For example, most participants expressed sexuality, PrEP, and HIV stigma as barriers to PrEP adherence and social wall of P3T provides an opportunity for positive social interaction and support related to these issues to take place, which has been shown in previous studies to be associated with reduced stigma [50,51].

However, the study participants provided critical recommendations for P3T app adaptations in the way the core features were presented that address cultural and contextual factors specifically relevant to Thai YMSM, which are expected to substantially enhance app acceptability, relevance, user engagement, and utility. For example, Thai YMSM voiced a strong preference for more visual content, such as infographics and video content, as opposed to text, similar to findings from formative work to develop P3 [34]. This may be because of the interest in consuming information via different mediums [35,36,52-54]. Therefore, these factors highlight the importance of understanding the reading ability and reading norms of the target population when developing app interventions [55]. The desire to keep the app fun and light through the use of stickers, a lighter color scheme, and ongoing seasonal promotions and games also featured prominently in both FGDs and KIIs. These suggestions reflect Thai culture and thus emphasize the importance of understanding cultural preferences and popular marketing practices when adapting mHealth interventions for use in different countries [56-58].

Both YMSM and key informants noted the strong influence peers have on the health beliefs, decision-making processes, and behaviors of Thai YMSM. This phenomenon has been identified in research in other cultures [59,60], and a study of Thai YMSM found that health beliefs and self-efficacy were associated with HIV prevention behaviors [52]—factors known to be influenced by peer groups [61]—were associated with HIV prevention behaviors [52]. App interventions focused on HIV prevention for Thai YMSM should assess whether functions that acknowledge the importance of peer relationships increase the likelihood of achieving the desired behavior. In P3T, peer interactions are possible via the social wall, and positive peer norms are presented in the videos added to the knowledge center that feature peer influencers. Future development of P3T will include the ability to visually represent the popularity of content among users via functions such as number of likes on comments and most popular topics, which could further capitalize on the influence of peers to promote behavior change.

There have been a small number of mobile apps that have tested YMSM at risk for HIV to support PrEP adherence and HIV prevention, mostly based in the United States. The availability of having reliable information on HIV and sexual health, reminder features for users to access relevant health care such as drug adherence and medical appointments, esthetically pleasing features for target users, and ease of use are common findings with what was found in this study from multiple studies previously conducted, including the LYNX and MyChoices apps [62-64]. Similar needs for medication reminders and education on their condition have been found in adolescents with other chronic health conditions, such as sickle cell disease [65]. No striking contrasts in YMSM needs using PrEP adherence apps in other studies compared with findings from those found in our study are apparent so far; however, many studies are underway that may produce differing outcomes to what has been seen so far [62,63,66]. Given the future potential scalability of the use of mobile phone apps in supporting PrEP adherence, understanding differences in feature preferences in different contexts, such as with age, gender identity, urban versus rural dwellers, and specific user profiles are valuable in adjusting mobile health technologies to maximize their benefits to specific populations [67]. Given the time limitations with data collection in adapting mobile technologies to user needs, future studies may benefit from sequential adaptive research trial designs to address this challenge.

Future Directions

We identified the benefits and challenges of adapting P3 for YMSM in Thailand. First, the remarkable similarities in PrEP adherence facilitators and barriers among YMSM in our study,

and in other studies, along with the general acceptability of the key features of P3 among our study participants support the growing trend of adapting evidenced-based apps, such as P3, for different cultures and contexts rather than developing new apps from scratch [31]. However, extensive formative research is critical to ensure that an app is appropriately adapted to meet the unique needs and population preferences in different settings [67,68]. Adapting an existing app with a common behavior change goal can increase the speed of app development, reduce costs, and ultimately expedite scalability [31-33].

However, even minor app changes or the addition of a few new features can require significant time and budgetary resources. Planning for a budget that is sufficient for formative research and app adaptations may ensure that the adapted app is culturally and contextually relevant and may shorten the time between adaptation, pilot study, and, if feasible and acceptable, a larger trial. Owing to the limited time and funding of a pilot grant and working with a commercial app developer based in North America, we were unable to implement all design changes that would better align the app with cultural preferences identified in the FGDs and KIIs. These changes focused on stylistic preferences and the inclusion of additional functions within the existing features. The main changes suggested included making the app color scheme lighter, offering customizable skins, adding stickers to enhance communication in the social wall, adding an app search function, adding engagement statistics to visualize content popularity among peers (eg, number of times a knowledge center article or social wall post has been read), adding health behavior tracking features that would allow for determination of HIV risk level based on data entered into the app, and regular content updates to make the app more dynamic. An additional need identified in using an app such as P3T in Thailand is the ability of the more features of the app to function offline. According to the study participants, most Thai YMSM are unable to afford data plans that would allow for consistent and reliable internet access. To maximize access of users to P3T, more extensive offline functionality is critically important in the Thai context. If justified by the findings from the pilot study, the remaining adaptations, along with additional findings anticipated from the pilot study, are planned for a subsequent development phase that could also inform the adaptation of other apps.

Including local designers and developers on the team may help stretch limited resources and ensure that local expertise leads to the cultural adaptation process [68]. For example, through the support of a Thai collaborator skilled in graphic design, we were able to make important design changes to the app that reflected a rich understanding of the esthetic preferences of Thai YMSM. The local designer rapidly developed avatars that were highly responsive to study participant preferences. We also faced unforeseen formatting challenges translating the content of the app from English to Thai, which substantially increased the amount of time required for this stage. Having a fluent Thai speaker on the development team could help address these and similar barriers in the adaptation process. Local app designers and developers also provide a capacity-building opportunity for local collaborators, which can then reduce cost and time during both adaptation and scale-up phases. Given that, particularly in the context of limited resources such as where this study was performed, it must be acknowledged that there are limited data to support the cost-effectiveness of digital health interventions [69]. Such considerations are important when considering future scalability and sustainability to inform policies on investments in this area of health care interventions [70].

In the COVID-19 era, it must also be noted that with its considerable challenges, there are opportunities to optimize digital approaches as well as psychosocial intervention health care delivery in adolescents. Multiple areas regarding this with relevance to future studies related to the P3T app could be explored, including the impact of the app on preventive behaviors, medication adherence, self-management, and cost-effectiveness [71,72].

Strengths

A major strength of this study was that it addressed a knowledge gap in the feasibility and acceptability of mHealth in low-middle income settings [29] and also in adolescent populations [28]. In addition, it provided practical guidance for adaptation of an existing mHealth platform in low-middle income settings rather than starting from scratch, a practical solution in the context of limited financial and staff resources as well as difficult-to-reach populations. It also highlighted the importance of using local expertise and capacity building in producing well-suited and sustainable interventions where mHealth interventions are adapted to very different settings for which they were originally intended. Finally, this study involved target user input from the start of its adaptation to the Thai context, an approach known to improve target user short- and long-term engagement [65,73-75].

Limitations

This study was conducted only in Bangkok; therefore, data collected from study participants may not reflect the perspectives of YMSM and PrEP providers nationwide. In addition, the majority of FGD participants were enrolled in full-time education, which in the context of Thailand signifies a higher socioeconomic status. Therefore, these findings may not be generalizable to the Thai YMSM of all socioeconomic groups. As parental consent was required for FGD participants under the age of 18 years, we were unable to enroll participants in this age group because of the unwillingness of potential participants to reveal their sexual orientation or risk behaviors to their parents. This may have led to findings biased toward the views, needs, and developmental requirements of older adolescents. In most countries, formidable challenges remain in conducting research with those under the age of 18 years, despite the need to gather information to inform interventions that address the current global HIV epidemic, which disproportionately affects young people [5,76]. In Thailand, parental consent is exempted for individuals under 18 years of age to receive sexually transmitted infection care, HIV testing, or PrEP initiation. However, the decision to allow those under 18 years to participate in HIV prevention research without parental consent is at the discretion of the ethical review boards. On the basis of our experience, some review boards in Thailand are willing to allow exemption of parental consent for adolescents in HIV

prevention trials where study involvement is a direct benefit to participants from this key population.

Despite these limitations, the adapted P3T app represents a significant step toward expansion and scale up of an app-based intervention to support PrEP adherence in Thailand, where PrEP has recently become available free of charge for YMSM through a national health coverage [5,77]. Additional studies in different geographical areas of Thailand with a high YMSM population density and studies focusing on young transgender women who have sex with men and other key populations would help further the adaptation processes of mHealth apps for those disproportionately affected by HIV in Thailand.

Conclusions

As observed in this study, most facilitators and barriers for PrEP adherence in the Thai context are similar to those observed in other global settings. The core features of the P3T app address the main facilitators and barriers related to PrEP adherence in Thailand for YMSM. The majority of the improvements suggested in this study focus on customizing the app to suit the stylistic presentation preferences of Thai YMSM and tailoring the app content to suit the cultural context of Thailand and the cultural nuances of the YMSM demography. Given that many of the facilitators and barriers are similar globally, adapting existing successful mHealth solutions will be the most effective and cost-efficient way forward as opposed to developing new apps. However, to obtain optimal results, it is vital that local app developers are included as partners in the adaptation process.

Acknowledgments

This study was funded by the National Institutes of Health, Fogarty International Center and the Ratchadapisek Sompoch Endowment Fund (2019) under the Telehealth Cluster, Chulalongkorn University.

Conflicts of Interest

None declared.

References

1. UNAIDS data 2019. UNAIDS. 2019 Dec 4. URL: https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data [accessed 2019-02-27]
2. van Griensven F, Thienkrua W, McNicholl J, Wimonsate W, Chaikummao S, Chonwattana W, et al. Evidence of an explosive epidemic of HIV infection in a cohort of men who have sex with men in Thailand. AIDS 2013 Mar 13;27(5):825-832. [doi: 10.1097/QAD.0b013e32835c546e] [Medline: 23169330]
3. van Griensven F, Holtz TH, Thienkrua W, Chonwattana W, Wimonsate W, Chaikummao S, et al. Temporal trends in HIV-1 incidence and risk behaviours in men who have sex with men in Bangkok, Thailand, 2006-13: an observational study. Lancet HIV 2015 Feb;2(2):64-70. [doi: 10.1016/S2352-3018(14)00031-9] [Medline: 26424462]
4. Thienkrua W, van Griensven F, Mock PA, Dunne EF, Raengsakulrach B, Wimonsate W, et al. Young men who have sex with men at high risk for HIV, Bangkok MSM cohort study, Thailand 2006-2014. AIDS Behav 2018 Jul;22(7):2137-2146 [FREE Full text] [doi: 10.1007/s10461-017-1963-7] [Medline: 29138981]
5. Pettifor A, Stoner M, Pike C, Bekker L. Adolescent lives matter: preventing HIV in adolescents. Curr Opin HIV AIDS 2018 May;13(3):265-273 [FREE Full text] [doi: 10.1097/COH.0000000000000453] [Medline: 29528850]
6. Grant R, Lama J, Anderson P, McMahan V, Liu A, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
7. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, iPrEx study team. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829 [FREE Full text] [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
8. Grant R, Anderson P, McMahan V, Liu A, Amico KR, Mehrotra M, for the iPrEx study team. Results of the iPrEx open-label extension (iPrEx OLE) in men and transgender women who have sex with men: PrEP uptake, sexual practices, and HIV incidence. In: Proceedings of the 20th International AIDS Conference (AIDS 2014). 2014 Presented at: 20th International AIDS Conference (AIDS 2014); July 20-25, 2014; Melbourne, Australia URL: https://www.mdlinx.com/internal-medicine/conference-abstract.cfm/18761/?nonus=0&searchstring=&coverage_day=0&page=1
9. Colby D, Srithanaviboonchai K, Vanichseni S, Ongwandee S, Phanuphak N, Martin M, et al. HIV pre-exposure prophylaxis and health and community systems in the Global South: Thailand case study. J Int AIDS Soc 2015 Jul 20;18(4 Suppl 3):19953 [FREE Full text] [doi: 10.7448/IAS.18.4.19953] [Medline: 26198342]
10. Essentials of HIV/AIDS treatment and prevention 2014. Bureau of AIDS, TB and STIs, Department of Disease Control, Ministry, of Public Health, Thailand. URL: https://www.rihes.cmu.ac.th/Ped_HIV/02-Guideline/1_Thai/2557/GL2557.pdf [accessed 2019-02-18]

11. Description news. National Health Security Office, Bangkok. URL: http://eng.nhso.go.th/view/1/Description_news/NewsAndEvents/4/EN-US [accessed 2020-03-13]
12. PrEP HIV-prevention med makes progress in Thailand. National News Bureau of Thailand. URL: http://thainews.prd.go.th/en/news/detail/TCATG191029161049997 [accessed 2020-03-13]
13. Phanuphak N, Sungsing T, Jantarapakde J, Pengnonyang S, Trachunthong D, Mingkwanrungruang P, et al. Princess PrEP program: the first key population-led model to deliver pre-exposure prophylaxis to key populations by key populations in Thailand. Sex Health 2018 Nov;15(6):542-555. [doi: 10.1071/SH18065] [Medline: 30249317]
14. Seekaew P, Nguyen E, Sungsing T, Jantarapakde J, Pengnonyang S, Trachunthong D, et al. Correlates of nonadherence to key population-led HIV pre-exposure prophylaxis services among Thai men who have sex with men and transgender women. BMC Public Health 2019 Mar 21;19(1):328-32- [FREE Full text] [doi: 10.1186/s12889-019-6645-0] [Medline: 30898095]
15. Wheeler DP, Fields SD, Beauchamp G, Chen YQ, Emel LM, Hightow-Weidman L, et al. Pre-exposure prophylaxis initiation and adherence among Black men who have sex with men (MSM) in three US cities: results from the HPTN 073 study. J Int AIDS Soc 2019 Feb;22(2):e25223 [FREE Full text] [doi: 10.1002/jia2.25223] [Medline: 30768776]
16. Pasipanodya EC, Jain S, Sun X, Blumenthal J, Ellorin E, Corado K, California Collaborative Treatment Group (CCTG) TAPIR Study Team. Trajectories and predictors of longitudinal preexposure prophylaxis adherence among men who have sex with men. J Infect Dis 2018 Oct 05;218(10):1551-1559 [FREE Full text] [doi: 10.1093/infdis/jiy368] [Medline: 30295803]
17. The 2018 household survey on the use of information and communication technology (quarter 1). National Statistical Office. URL: https://tinyurl.com/n53davf3 [accessed 2021-08-11]
18. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):186-194 [FREE Full text] [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
19. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
20. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
21. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
22. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
23. Thakkar J, Kurup R, Laba TL, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
24. Phanuphak N, Anand T, Jantarapakde J, Nitpolprasert C, Himmad K, Sungsing T, et al. What would you choose: online or offline or mixed services? Feasibility of online HIV counselling and testing among Thai men who have sex with men and transgender women and factors associated with service uptake. J Int AIDS Soc 2018 Jul;21(Suppl 5):e25118 [FREE Full text] [doi: 10.1002/jia2.25118] [Medline: 30033644]
25. Sullivan PS, Jones J, Kishore N, Stephenson R. The roles of technology in primary hiv prevention for men who have sex with men. Curr HIV/AIDS Rep 2015 Dec;12(4):481-488. [doi: 10.1007/s11904-015-0293-5] [Medline: 26519083]
26. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
27. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, technology, and HIV: recent advances and future directions. Curr HIV/AIDS Rep 2015 Dec;12(4):500-515 [FREE Full text] [doi: 10.1007/s11904-015-0280-x] [Medline: 26385582]
28. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
29. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
30. Guadamuz T, Cheung D, Wei C, Koe S, Lim S. Young, online and in the dark: scaling up HIV testing among MSM in ASEAN. PLoS One 2015 May 14;10(5):e0126658 [FREE Full text] [doi: 10.1371/journal.pone.0126658] [Medline: 25973907]
31. Mulawa MI, LeGrand S, Hightow-Weidman LB. eHealth to enhance treatment adherence among youth living with HIV. Curr HIV/AIDS Rep 2018 Aug;15(4):336-349 [FREE Full text] [doi: 10.1007/s11904-018-0407-y] [Medline: 29959649]

32. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
33. Hightow-Weidman LB, Muessig K, Rosenberg E, Sanchez T, LeGrand S, Gravens L, et al. University of North Carolina/Emory Center for Innovative Technology (iTech) for addressing the HIV epidemic among adolescents and young adults in the United States: protocol and rationale for center development. JMIR Res Protoc 2018 Aug 03;7(8):e10365 [FREE Full text] [doi: 10.2196/10365] [Medline: 30076126]
34. LeGrand S, Knudtson K, Benkeser D, Muessig K, Mcgee A, Sullivan PS, et al. Testing the efficacy of a social networking gamification app to improve pre-exposure prophylaxis adherence (P3: Prepared, Protected, emPowered): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e10448 [FREE Full text] [doi: 10.2196/10448] [Medline: 30563818]
35. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
36. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
37. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
38. NVivo qualitative data analysis software. QSR International Pty Ltd. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-08-11]
39. Statistics and facts about LINE. Statista. URL: https://www.statista.com/topics/1999/line/ [accessed 2020-04-28]
40. Nguyen LH, Tran BX, Rocha LE, Nguyen HL, Yang C, Latkin CA, et al. A systematic review of eHealth interventions addressing HIV/STI prevention among men who have sex with men. AIDS Behav 2019 Sep;23(9):2253-2272 [FREE Full text] [doi: 10.1007/s10461-019-02626-1] [Medline: 31401741]
41. Chemnasiri T, Varangrat A, Amico KR, Chitwarakorn A, Dye BJ, Grant RM, HPTN 067/ADAPT Study Team. Facilitators and barriers affecting PrEP adherence among Thai men who have sex with men (MSM) in the HPTN 067/ADAPT Study. AIDS Care 2020 Feb;32(2):249-254 [FREE Full text] [doi: 10.1080/09540121.2019.1623374] [Medline: 31159584]
42. Taylor SW, Mayer KH, Elsesser SM, Mimiaga MJ, O'Cleirigh C, Safren SA. Optimizing content for pre-exposure prophylaxis (PrEP) counseling for men who have sex with men: perspectives of PrEP users and high-risk PrEP naïve men. AIDS Behav 2014 May;18(5):871-879 [FREE Full text] [doi: 10.1007/s10461-013-0617-7] [Medline: 24077928]
43. Peng P, Su S, Fairley CK, Chu M, Jiang S, Zhuang X, et al. A global estimate of the acceptability of pre-exposure prophylaxis for HIV among men who have sex with men: a systematic review and meta-analysis. AIDS Behav 2018 Apr;22(4):1063-1074. [doi: 10.1007/s10461-017-1675-z] [Medline: 28176168]
44. Tangmunkongvorakul A, Chariyalertsak S, Amico KR, Saokhieo P, Wannalak V, Sangangamsakun T, et al. Facilitators and barriers to medication adherence in an HIV prevention study among men who have sex with men in the iPrEx study in Chiang Mai, Thailand. AIDS Care 2013 Aug;25(8):961-967. [doi: 10.1080/09540121.2012.748871] [Medline: 23252473]
45. Wheelock A, Eisingerich AB, Ananworanich J, Gomez GB, Hallett TB, Dybul MR, et al. Are Thai MSM willing to take PrEP for HIV prevention? An analysis of attitudes, preferences and acceptance. PLoS One 2013;8(1):e54288 [FREE Full text] [doi: 10.1371/journal.pone.0054288] [Medline: 23342121]
46. Gilmore HJ, Liu A, Koester KA, Amico KR, McMahan V, Goicochea P, et al. Participant experiences and facilitators and barriers to pill use among men who have sex with men in the iPrEx pre-exposure prophylaxis trial in San Francisco. AIDS Patient Care STDS 2013 Oct;27(10):560-566 [FREE Full text] [doi: 10.1089/apc.2013.0116] [Medline: 24093809]
47. Hosek S, Siberry G, Bell M, Lally M, Kapogiannis B, Green K, Adolescent Trials Network for HIVAIDS Interventions (ATN). The acceptability and feasibility of an HIV preexposure prophylaxis (PrEP) trial with young men who have sex with men. J Acquir Immune Defic Syndr 2013 Apr 01;62(4):447-456 [FREE Full text] [doi: 10.1097/QAI.0b013e3182801081] [Medline: 24135734]
48. Zablotska I, Grulich AE, Phanuphak N, Anand T, Janyam S, Poonkasetwattana M, et al. PrEP implementation in the Asia-Pacific region: opportunities, implementation and barriers. J Int AIDS Soc 2016 Oct 18;19(7(Suppl 6)):21119 [FREE Full text] [doi: 10.7448/IAS.19.7.21119] [Medline: 27760688]
49. Kebaabetswe PM, Stirratt MJ, McLellan-Lemal E, Henderson FL, Gray SC, Rose CE, et al. Factors associated with adherence and concordance between measurement strategies in an HIV daily oral tenofovir/emtricitibine as Pre-exposure Prophylaxis (Prep) clinical trial, Botswana, 2007-2010. AIDS Behav 2015 May;19(5):758-769 [FREE Full text] [doi: 10.1007/s10461-014-0891-z] [Medline: 25186785]
50. Bauermeister JA, Muessig KE, LeGrand S, Flores DD, Choi SK, Dong W, et al. Hiv and sexuality stigma reduction through engagement in online forums: results from the HealthMPowerment intervention. AIDS Behav 2019 Mar;23(3):742-752 [FREE Full text] [doi: 10.1007/s10461-018-2256-5] [Medline: 30121727]
51. Whiteley L, Mena L, Craker LK, Healy MG, Brown LK. Creating a theoretically grounded gaming app to increase adherence to pre-exposure prophylaxis: lessons from the development of the viral combat mobile phone game. JMIR Serious Games 2019 Mar 27;7(1):e11861 [FREE Full text] [doi: 10.2196/11861] [Medline: 30916652]

52. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
53. Human development reports. United Nations Development Programme. URL: http://hdr.undp.org/en/countries/profiles/THA [accessed 2020-04-29]
54. Factors affecting the promotion of a reading culture in Thailand. The Thai Health Promotion Foundation (ThaiHealth). 2011. URL: https://www.tkpark.or.th/stocks/extra/00052d.pdf [accessed 2020-05-06]
55. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of flesch formula. Educ Health (Abingdon) 2017;30(1):84-88 [FREE Full text] [doi: 10.4103/1357-6283.210517] [Medline: 28707643]
56. Ben-Zeev D. Mobile health for mental health in West Africa: the case for Ghana. Psychiatr Serv 2018 Jul 01;69(7):741-743. [doi: 10.1176/appi.ps.201700555] [Medline: 29540120]
57. Maglalang DD, Yoo GJ, Ursua RA, Villanueva C, Chesla CA, Bender MS. "I don't have to explain, people understand": acceptability and cultural relevance of a mobile health lifestyle intervention for Filipinos with type 2 diabetes. Ethn Dis 2017 Apr 20;27(2):143-154 [FREE Full text] [doi: 10.18865/ed.27.2.143] [Medline: 28439185]
58. McCool J, Tanielu H, Umali E, Whittaker R. Assessing the cross-cultural adaptation and translation of a text-based mobile smoking cessation program in Samoa (TXTTaofiTapaa): pilot study. JMIR Mhealth Uhealth 2018 Aug 31;6(8):e173 [FREE Full text] [doi: 10.2196/mhealth.9033] [Medline: 30170994]
59. Asrese K, Mekonnen A. Social network correlates of risky sexual behavior among adolescents in Bahir Dar and Mecha Districts, North West Ethiopia: an institution-based study. Reprod Health 2018 Apr 11;15(1):61 [FREE Full text] [doi: 10.1186/s12978-018-0505-8] [Medline: 29642938]
60. Widman L, Choukas-Bradley S, Helms SW, Prinstein MJ. Adolescent susceptibility to peer influence in sexual situations. J Adolesc Health 2016 Mar;58(3):323-329 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.10.253] [Medline: 26794431]
61. Tomé G, Matos M, Simões C, Diniz JA, Camacho I. How can peer group influence the behavior of adolescents: explanatory model. Glob J Health Sci 2012 Feb 29;4(2):26-35 [FREE Full text] [doi: 10.5539/gjhs.v4n2p26] [Medline: 22980148]
62. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
63. Liu A, Coleman K, Bojan K, Serrano PA, Oyedele T, Garcia A, et al. Developing a mobile app (LYNX) to support linkage to HIV/sexually transmitted infection testing and pre-exposure prophylaxis for young men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659 [FREE Full text] [doi: 10.2196/10659] [Medline: 30681964]
64. Biello KB, Hill-Rorie J, Valente PK, Futterman D, Sullivan PS, Hightow-Weidman L, et al. Development and evaluation of a mobile app designed to increase HIV testing and pre-exposure prophylaxis use among young men who have sex with men in the United States: open pilot trial. J Med Internet Res 2021 Mar 24;23(3):e25107 [FREE Full text] [doi: 10.2196/25107] [Medline: 33759792]
65. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
66. Hightow-Weidman L, Muessig K, Claude K, Roberts J, Zlotorzynska M, Sanchez T. Maximizing digital interventions for youth in the midst of COVID-19: lessons from the adolescent trials network for HIV interventions. AIDS Behav 2020 Aug;24(8):2239-2243 [FREE Full text] [doi: 10.1007/s10461-020-02870-w] [Medline: 32306214]
67. Chávez NR, Shearer LS, Rosenthal SL. Use of digital media technology for primary prevention of STIs/HIV in youth. J Pediatr Adolesc Gynecol 2014 Oct;27(5):244-257. [doi: 10.1016/j.jpag.2013.07.008] [Medline: 24332613]
68. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Design, development, and formative evaluation of a smartphone application for recording and monitoring physical activity levels: the 10,000 Steps "iStepLog". Health Educ Behav 2013 Apr;40(2):140-151. [doi: 10.1177/1090198112449460] [Medline: 22984196]
69. Stephan LS, Almeida ED, Guimaraes RB, Ley AG, Mathias RG, Assis MV, et al. Processes and recommendations for creating mHealth apps for low-income populations. JMIR Mhealth Uhealth 2017 Apr 03;5(4):e41 [FREE Full text] [doi: 10.2196/mhealth.6510] [Medline: 28373155]
70. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb 2;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
71. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121. [doi: 10.2196/mhealth.6425] [Medline: 27780795]
72. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]

73. Serlachius A, Badawy SM, Thabrew H. Psychosocial challenges and opportunities for youth with chronic health conditions during the COVID-19 pandemic. JMIR Pediatr Parent 2020 Oct 12;3(2):e23057 [FREE Full text] [doi: 10.2196/23057] [Medline: 33001834]
74. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
75. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
76. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
77. Parental consent undermines the right to health of adolescents. UNAIDS. 2020. URL: https://www.unaids.org/en/resources/presscentre/featurestories/2020/march/20200316_parental-consent [accessed 2020-03-06]

Abbreviations

FGD: focus group discussion
KII: key informant interview
mHealth: mobile health
MSM: men who have sex with men
P3: Prepared, Protected, emPowered
P3T: P3 Thailand
PrEP: pre-exposure prophylaxis
TRCARC: Thai Red Cross AIDS Research Center
VCT: voluntary counseling and testing
YMSM: young men who have sex with men

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by C Lelutiu-Weinberger, S Badawy, MD, MS; comments to author 11.02.21; revised version received 05.07.21; accepted 21.07.21; published 04.11.21.

Please cite as:
Songtaweesin WN, LeGrand S, Bandara S, Piccone C, Wongharn P, Moonwong J, Jupimai T, Saisaengjan C, Theerawit T, Muessig K, Hightow-Weidman L, Puthanakit T, Phanuphak N, Tangmunkongvorakul A
Adaptation of a Theory-Based Social Networking and Gamified App-Based Intervention to Improve Pre-Exposure Prophylaxis Adherence Among Young Men Who Have Sex With Men in Bangkok, Thailand: Qualitative Study
J Med Internet Res 2021;23(11):e23852
URL: https://www.jmir.org/2021/11/e23852
doi:10.2196/23852
PMID:34734828

Original Paper

A Digital Self-management Program (Help to Overcome Problems Effectively) for People Living With Cancer: Feasibility Randomized Controlled Trial

Hayley Wright[1*], BSc, MSc, PhD; Faith Martin[1*], PhD, DClinPsy; Wendy Clyne[2*], BA, PhD; Cain C T Clark[1*], BSc, MSc, PhD; Gabriela Matouskova[3*], MBA; Michael McGillion[4*], BScN, PhD; Andrew Turner[1*], BSc, PhD

[1]Centre for Intelligent Healthcare, Research Institute for Health and Wellbeing, Coventry University, Coventry, United Kingdom

[2]National Institute for Health Research, Research Design Service South West, Peninsula Medical School, University of Plymouth, Devon, United Kingdom

[3]Hope For The Community, Community Interest Company, The Enterprise Hub, Coventry, United Kingdom

[4]School of Nursing, McMaster University, Hamilton, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Andrew Turner, BSc, PhD
Centre for Intelligent Healthcare
Research Institute for Health and Wellbeing
Coventry University
Priory Street
Coventry, CV1 5FB
United Kingdom
Phone: 44 7599782465
Email: hsx116@coventry.ac.uk

Abstract

Background: We present the results of a feasibility, randomized waitlist control group (CG) parallel design study with a 1:1 allocation ratio. Participants were randomized into an intervention group (IG) or a waitlist CG. The intervention was a 6-week digital self-management program, Help to Overcome Problems Effectively (HOPE), for people with cancer.

Objective: This study aims to test the feasibility of a digitally delivered self-management program for people with cancer. This will inform the design of a definitive randomized controlled trial. In addition, a preliminary assessment of the impact of the HOPE program via secondary outcomes will be used to assess signals of efficacy in a trial context.

Methods: Participants were drawn from an opportunity sample, referred by Macmillan Cancer Support, and were invited via email to participate in the study (N=61). Primary outcomes were rates of recruitment, retention, follow-up, completion and adherence, sample size and effect size estimation, and assessment of progression criteria for a definitive trial. Secondary outcomes were self-report measures of participants' positive mental well-being, depression, anxiety, and patient activation (ie, confidence in managing their cancer). The intervention and data collection took place on the web.

Results: The recruitment rate was 77% (47/61). A total of 41 participants completed the baseline questionnaires and were randomized to either the IG (n=21) or the waitlist CG (n=20). The retention rate (attending all program sessions) was greater than 50% (all: 21/41, 51%, IG: 10/21, 48%; and CG: 11/20, 55%). The follow-up rate (completing all questionnaires) was greater than 80% (all: 33/41, 80%; IG: 16/21, 76%; and CG: 17/20, 85%). The completion rate (attending ≥3 sessions and completing all questionnaires) was greater than 60% (all: 25/41, 61%; IG: 13/21, 62%; and CG: 12/20, 60%). Engagement data showed that participants viewed between half (5.1/10, 51%) and three-quarters (12.2/16, 76%) of the pages in each session.

Conclusions: All progression criteria for a definitive trial were met, as supported by the primary outcome data. The IG showed improved postprogram scores on measures of positive mental well-being, depression, anxiety, and patient activation. A full-scale trial of the digital HOPE program for people with cancer will allow us to fully evaluate the efficacy of the intervention relative to a CG.

Trial Registration: ISRCTN Registry ISRCTN79623250; http://www.isrctn.com/ISRCTN79623250

International Registered Report Identifier (IRRID): RR2-10.2196/24264

(J Med Internet Res 2021;23(11):e28322) doi:10.2196/28322

KEYWORDS

self-management; cancer; survivorship; digital; positive psychology

Introduction

Background and Rationale

As of May 2021, the United Kingdom has seen 3 periods of national lockdown and widespread social and physical distancing measures implemented by the government in an attempt to curtail the spread of SARS-CoV-2. Such measures have resulted in significant reductions in the delivery of cancer services in the United Kingdom, as pressures on health services, lockdown demands, and the need to reduce face-to-face interactions have taken precedence [1]. Remote emotional support for these patients has recently been recommended by researchers [2] and cancer support specialists [3]. People with cancer are now experiencing increased health anxiety [4] owing to the risk of serious complications from the virus for people with cancer [5,6]. The need to implement alternative models of care during COVID-19, including more self-management support, has been noted by leading cancer care experts [7]. Moreover, before the COVID-19 pandemic, the UK National Health Service (NHS) had already called for greater emphasis on the delivery of digital, holistic, person-centered care for people with cancer [8]. Effective digital interventions have never been a more important aspect of care.

People with cancer are already known to face multiple challenges in terms of mental and physical health following primary treatment, including fatigue, pain, sexual problems, cognitive functioning, depression, anxiety, social isolation, and financial issues [9-13]. Furthermore, in the longer term, many patients with cancer experience negative impacts on their psychological well-being and mental health, including hypervigilance, anxiety, posttraumatic stress, and depression [14-20]. Research shows that 2 years after diagnosis, up to 20% of cancer survivors met the criteria for major depression and up to 40% met the criteria for an anxiety disorder [18-20]. As a result of the COVID-19 pandemic, people with cancer have reported a further decline in their mental and physical health [3].

Before the COVID-19 pandemic, there was a shortage of accessible self-management interventions, and there was an even greater need for digital interventions to comply with social distancing guidelines. Around a decade ago, we co-designed a face-to-face self-management program for survivors of all types of cancer [21,22]. People with cancer, oncologists, specialist cancer nurses, and representatives from a leading UK cancer charity (Macmillan Cancer Support; MCS) were involved in the co-design process. This led to the development of the *Help to Overcome Problems Effectively* (HOPE) program, which has been described in detail elsewhere [23,24]. The HOPE program aims to enhance well-being by fostering positive emotions and stimulating positive functioning. A parallel goal is to reduce depressive symptoms. The HOPE program is based on principles derived from positive psychology and focuses on positive experiences, strengths, and personal competencies rather than mental health problems such as anxiety and depression. It incorporates evidence-based exercises based on positive psychology, in addition to elements stemming from mindfulness, cognitive behavioral therapy, and problem-solving therapy. The HOPE program recognizes the common challenges and unmet needs across all types of cancer, including fatigue, fear of recurrence, and psychological distress [9-20]. The HOPE program was designed to provide support for these most common, typically overlapping needs in people living with most types of cancer. We regularly consult with MCS on their eHealth Needs Assessment data and review the most common needs indicated by people living with all types of cancer. The HOPE program provides general psychological and well-being support based on these needs and is open to all adult cancer survivors. The HOPE program differs from many other cancer self-management programs due to the focus on (1) positive psychology [25-27], (2) hope and gratitude [28] to improve well-being and coping, (3) co-created content, and (4) peer-facilitated delivery. The HOPE program is moderated by trained peer facilitators who are affected by cancer in some way. The facilitators received training from MCS and followed a delivery protocol. The facilitator's role is to offer encouragement to participants and stimulate discussion in social networking forums by inviting participants to respond with comments to specific questions or respond to questions or comments posted by participants. Facilitators also monitor daily social networking posts for safety and report technical problems to the research team. The facilitators spent approximately 2 hours per session, supporting the participants. The in-person program was adapted for digital delivery (see [24] for full details of adaptation), using a user-centered, iterative approach [29]. A set of design requirements and a design brief were drawn up in consultation with end users and stakeholders. The initial digital version of HOPE went through a number of iterative testing sessions, with improvements made to usability after each iteration. These iterations were intended to develop a system that was usable and accepted by the intended user group to increase the likelihood of uptake and continued use and to ensure that the technology did not prove a barrier to engagement and participation. Initial evaluation has suggested positive effects on anxiety, depression, and positive well-being in people with cancer, with positive user feedback [23]. This suggests that a trial of the digital HOPE program might be viable and meaningful. A feasibility randomized controlled trial (RCT) study of the digital intervention was required to assess whether participants consent to be randomized and to test the feasibility of running a wait-list control study design of the HOPE program.

Objectives

This study aims to test the feasibility of a digitally delivered self-management program for people with cancer. This will inform the design of a definitive RCT. In addition, a preliminary assessment of the impact of the HOPE program via secondary

outcomes will be used to assess signals of efficacy in a trial context.

The planned primary outcomes (trial feasibility objectives) of the study were as follows: (1) recruitment rates for participation and for randomization; (2) retention and follow-up rates as the participants move through the trial; (3) adherence rates to study procedures, intervention attendance, and engagement; (4) sample size and effect size estimation for a definitive trial; and (5) progression criteria for a definitive trial.

The secondary outcomes related to participant well-being are measures of positive mental well-being, depression, anxiety, and confidence in self-managing cancer (patient activation), as indicated by scores on validated measures.

Methods

The following sections were written in accordance with the 2016 CONSORT (Consolidated Standards of Reporting Trials) extension for pilot and feasibility trials [30].

Trial Design

This study used a feasibility, randomized waitlist control group (CG) parallel design, with a 1:1 allocation ratio. Participants were randomized into an intervention group (IG) or a waitlist CG. The IG received access to the digital 6-week HOPE program immediately. The CG was placed on a waiting list for approximately 6 weeks, after which time they also received access to the same digital 6-week HOPE program. Key outcome measures were collected via web-based questionnaires at time 0 (T0; baseline) and time 1 (T1; 6 weeks postrandomization and postprogram for IG). We also sent the questionnaires to the CG only again after they had received the intervention (time 2; T2; postprogram for CG).

Participants

The participants were referred by MCS, a leading UK cancer charity. They advertise the HOPE program through their social media networks, MCS websites, and word-of-mouth through specialist nurses.

Eligibility criteria for participants were as follows: (1) any cancer diagnosis, at any treatment stage; (2) adult (18 years or older), (3) located in the United Kingdom; (4) access to the internet and a device that allows them to engage with the intervention; and (5) fluent in English to be able to engage with all the material in the intervention.

All study data were collected on the web via questionnaires administered through the Qualtrics Survey Software (Qualtrics 2019; [31]).

Intervention

All participants had access to the same digital HOPE program. The IG received access immediately, and the CG was granted access approximately 6 weeks later.

The HOPE program was delivered on the web. Full details of the digital HOPE program development, content, and weekly topics have been described elsewhere (see [23,24]), but we provide a brief overview here. All the HOPE program modules have the same structure and format, with a variety of components each week, focusing on a particular issue or a set of techniques, and ending with goal setting activity. The HOPE program is asynchronous, and content is released on a weekly basis at set times (eg, at midday every Monday) over the 6 weeks of the intervention. Forums and messaging facilities acted as a conduit for communication between participants and facilitators, and the program was moderated by trained peer facilitators. Table 1 provides an overview of the content and activities within each weekly module of the HOPE program.

Table 1. Examples of content, exercises, and activities within each weekly module of the Help to Overcome Problems Effectively program.

Session	Examples of content	Examples of exercises and activities (self-management tools)
Week 1 (Introduction or instilling hope)	• Aims of the program • User guide to navigating the platform and setting up a profile • Introduction to self-management • The benefits of positive emotions • Video (Positive emotions for a flourishing life) • The power of gratitude • Personalized goal-setting • Video (How to set achievable goals) • Forum topic (Reasons for joining the program) • Further resources and links (eg, videos, podcasts, and websites) to gratitude, positivity, and goal-setting	• Interactive gratitude diary • SMARTER[a] goal-setting • Assessment: positivity ratio test and positive and negative emotions test
Week 2 (Stress management)	• Understanding stress • Managing stress • Videos (How to manage stress and how to make stress your friend) • Coping with unhelpful thinking patterns • Mindfulness for stress management and meditation • Self-compassion and acceptance • Video (How to be kind to yourself) • Forum topic (How do you deal with cancer-related stress?) • Further resources and links (eg, videos, podcasts, and websites) to self-compassion, mindfulness, and stress management	• Interactive gratitude diary • SMARTER goal-setting and goal feedback • Guided relaxation and meditation exercise (podcasts) • How to cope with unhelpful thoughts (worksheet)
Week 3 (Managing fatigue)	• Understanding the boom and bust cycle • Using the 3 Ps (prioritizing, planning, and pacing) for managing fatigue • Video (Tips for managing fatigue) • Sleeping better; podcast: Tips to improve sleep • Forum topic (Coping with fatigue) • Further resources and links (eg, videos, podcasts, and websites) to sleeping better	• Interactive gratitude diary • SMARTER goal-setting and goal feedback • Fatigue and pacing diaries (worksheets) • Quiz (What are the main challenges faced by cancer survivors?)
Week 4 (Body image and communication)	• Body image • Video (Body image and cancer) • Sexuality and intimacy • Video (Cancer as a passport to emotional intimacy) • Communication skills and tips for talking with the health care team and family • Forum topic: experiences of coping with body changes and experiences of communicating with the health care team • Further resources and links (eg, videos, podcasts, and websites) to sexuality, intimacy, and relationships	• Interactive gratitude diary • SMARTER goal-setting and goal feedback
Week 5 (Physical activity and fear of recurrence)	• Coping with fear of recurrence • Videos (Moving forward while being worried about cancer returning and the regrets of those who are dying) • Hopes and dreams for the future • Video: Before I die project • The benefits of physical activity • Video (Tips for becoming and staying active) • Forum topic (Concerns about cancer coming back) • Further resources and links (eg, videos, podcasts, and websites) to managing concerns about cancer coming back and getting more active	• Interactive gratitude diary • SMARTER goal-setting and goal feedback

Session	Examples of content	Examples of exercises and activities (self-management tools)
Week 6 (Character strengths and happiness)	• Understanding how using your strengths can lead to a more fulfilling life • Video (The science of character strengths) • Tips for authentic happiness; managing setbacks and keeping going • Forum topic (Learning from the program) • Further resources and links (eg, videos, podcasts, and websites) to MCS[b] web-based communities and happiness resources	• Interactive gratitude diary • SMARTER goal-setting and goal feedback • Assessment (positivity ratio test and positive and negative emotions test and character strengths) • Quiz (What contributes to happiness?)

[a]SMARTER: SMARTER is an acronym used by many organizations for goal-setting, and stands for specific, measurable, achievable, relevant, time-bound, enjoyable, and reward.

[b]MCS: Macmillan Cancer Support.

Primary Outcomes

The primary outcome measures for this feasibility RCT were as follows:

Recruitment Rates

Recruitment rates for participation and randomization were collected primarily through Qualtrics at the start of the trial. All eligible participants identified by MCS were sent a link to the Qualtrics study survey. Recruitment rates were then calculated from the following: (1) providing consent and (2) completing baseline questionnaires. Direct email from participants indicating refusal or declining to participate in the study indicated a refusal. These participants were still offered access to the HOPE program but did not participate in any further data collection.

Retention, Follow-up, and Completion Rates

The participant retention rate was calculated as the percentage of participants attending all 6 program sessions. Studies show that a median of 56% of participants complete the full program in digital interventions for mental well-being [32,33]. As high rates of nonuse attrition [34] are common and of concern in digitally delivered interventions, and because of uncertainties relating to the COVID-19 pandemic, we set a more conservative target of 50% of participants completing all 6 sessions of the intervention.

Follow-up was calculated as the percentage of participants who completed all web-based study questionnaires. Participants who were lost to follow-up were identified through Qualtrics as those who did not complete the postprogram questionnaires at the end of the intervention period. It is possible that these participants may still have attended some portions or the entire HOPE program, despite not completing questionnaires.

If participants attended at least half of the intervention (3 sessions) [32] and completed the study questionnaires, they were classified as intervention completers. Studies show a nonlinear relationship between the time spent on an intervention, the number of sessions completed, and outcomes [32]. The amount of use needed to obtain desired outcomes varies across groups, and individuals may stop using the intervention once personal goals are achieved [35]. Therefore, we set a more pragmatic target for our primary outcome measure of *completion rate* of at least 3 sessions attended, and completion of all study questionnaires.

Adherence and Engagement Measures

The intervention platform collects user engagement data, such as the number of pages viewed in each session and the number of goals set that assists the moderators with participant engagement and experience. We measured the mean percentage of pages viewed per session, and the number of posts or comments a participant made for key activities (gratitude, setting goals, goal feedback, liking posts, and comments posted).

Sample Size and Effect Size Estimation for Future Definitive Trial

To inform the sample size estimation for a future definitive trial, we calculated the SDs of key continuous secondary outcomes at baseline. To estimate potential effect sizes for a primary outcome in a future definitive trial from pre- to postprogram, we calculated the difference between the mean difference pre- and postprogram for the IG and CG and divided by the pooled SD at baseline [36].

Progression Criteria

There is little guidance available for determining progression criteria for exploratory studies, including feasibility trials [37]. Following good practice, our progression criteria were discussed with Patient and Public Involvement representatives [38] and within the trial project group. Our 3 progression criteria reflect specific uncertainties regarding the feasibility of a larger definitive trial. The web-based HOPE program has not previously been delivered in an RCT study context, so two of our progression criteria tested the willingness of people to participate in a trial (recruitment and questionnaire completion rate). Full intervention completion (now labeled as “retention” within the manuscript) is the most frequently reported metric for adherence to web-based interventions [39], which, through discussion, was set as our third progression criteria.

To inform the progression to a definitive trial, we compared our results to the progression criteria set *a priori* as follows: (1) recruitment rate >70% of eligible participants consented, (2) questionnaire completion rate >70% of participants completing T1 questionnaires, and (3) retention rate >50% of participants attending all 6 HOPE program sessions.

Secondary Outcomes

Sociodemographic and health data were collected at T0. Participants were asked to provide the following information via a web-based questionnaire: gender, age, ethnicity, marital status, highest level of education, employment and occupation, and details about their cancer diagnosis and other medical conditions.

Participants completed a set of validated questionnaires at T0 (baseline), T1 (6 weeks postrandomization), and T2 (postprogram for CG only). Postprogram (T1 and T2) questionnaires were made available to participants the week after the intervention ended and remained available for a further 4 weeks. The positive mental well-being, depression, anxiety, and patient activation measures are detailed below.

The Warwick Edinburgh Mental Well-being Scale (WEMWBS) [40] is a validated scale of 14 positively worded feelings and thoughts used to assess mental well-being within the adult population. The scale includes measures of positive affect, satisfying interpersonal relationships and positive functioning, for example, *Below are some statements about feelings and thoughts. Please tick the box that best describes your experience of each over the last 2 weeks*...(1) "I have been feeling optimistic about the future," (2) "I have been thinking clearly," and (3) "I have been feeling loved." Participants rated each of the 14 items on a scale of 1 to 5 (1=none of the time, 2=rarely, 3=some of the time, 4=often, 5=all of the time), providing a total positive mental well-being score ranging from 14 to 70, with higher scores representing greater positive mental well-being. The WEMWBS had good internal consistency (Cronbach α=.91). A change of 3 or more was seen as clinically meaningful change [41].

The 9-item Patient Health Questionnaire (PHQ-9) [42] is a validated 9-item measure, which assesses the frequency of experience of the symptoms of depression. For example, *Over the past 2 weeks, how often have you been bothered by any of the following problems*...(1) *little interest or pleasure in doing things*; (2) *feeling down, depressed, or hopeless*; and (3) *poor appetite or overeating*.

Responses to each of the 9 items ranged from 0 to 3 (0=not at all, 1=several days, 2=more than half the days, 3=nearly every day), leading to a summed score between 0 and 27, with higher scores indicating greater severity of depression. The PHQ-9 has good internal consistency (α=.89). Scores of 10 or more are presumed to be above the clinical range; thus, participants scoring 10 are categorized as depressed for the purpose of this study.

The 7-item Generalized Anxiety Disorder scale (GAD-7) [43] is a validated 7-item scale measuring symptoms of generalized anxiety disorder, for example, *Over the past 2 weeks, how often have you been bothered by the following problems*...(1) *feeling nervous, anxious or on edge*; (2) *trouble relaxing*; and (3) *becoming easily annoyed or irritable*. Responses to all 7 items ranged from 0 to 3 (0=not at all, 1=several days, 2=more than half the days, 3=nearly every day), providing a total score of 0 to 21, with higher scores indicating greater anxiety. GAD-7 has good internal consistency (Cronbach =.92). Scores of 8 or more are presumed to be above the clinical range; therefore, participants scoring 8 are categorized as having anxiety for the purpose of this study.

The Patient Activation Measure [44] is a validated, licensed tool with good internal consistency (Cronbach =.81), which has been extensively tested with reviewed findings from a large number of studies. It helps to measure the spectrum of knowledge, skills, and confidence in patients and captures the extent to which people feel engaged and confident in taking care of their condition. Individuals are asked to complete a short survey, and based on their responses, they receive a Patient Activation Measure score (between 0 and 100). The resulting score places the individual at 1 of 4 levels of activation, each of which reveals insight into a range of health-related characteristics, including behaviors and outcomes. The 4 levels of activation are as follows:

1. Level 1 (scores ≤47.0): Individuals tend to be passive and feel overwhelmed by managing their own health. They may not understand their roles in the care process.
2. Level 2 (scores 47.1-55.1): Individuals may lack the knowledge and confidence to manage their health.
3. Level 3 (scores 55.2-67.0): Individuals appear to be taking action but may still lack the confidence and skill to support their behaviors.
4. Level 4 (scores 67.1): Individuals have adopted many of the behaviors needed to support their health but may not be able to maintain them in the face of life stressors.

Sample Size for This Study

All study participants were drawn from an opportunity sample (N=61), provided by MCS, of eligible candidates who expressed an interest in taking part in the HOPE program. An arbitrary sample size of n=40 was deemed adequate for this feasibility study, informed by similar studies in this area, with sample sizes ranging from 10 to 20 in each arm [45]. All potential study participants were emailed a link to the study website hosted by Qualtrics, where they were asked to read the digital Participant Information Sheet, read and agree to the statements on the digital consent form, and complete the digital T0 questionnaire before randomization.

Randomization

Sequence Generation

All participants who provided informed consent and completed the T0 questionnaires were randomized into the IG or CG using a 1:1 ratio via the randomization function within the Qualtrics Survey Software.

Allocation Concealment Mechanism

Participants were informed on completion of the T0 questionnaires, via a notification in Qualtrics, whether they had been randomized to the IG (in this case, starting in May 2020), or the CG (in this case, starting in June 2020). The research team remained unaware of participant allocation until group contact lists were created at the next data collection point (ie, T1).

Implementation

Participants were allocated to the IG or CG via the randomization function in Qualtrics. Participants were then emailed with a link to the HOPE program starting in the following week (IG), or a message to say that they would be emailed a link to the HOPE program (CG) in approximately 6 weeks' time.

Blinding

Owing to the nature of the study design, it was not possible to blind the participants to their group allocation. However, statistical analyses of study data were conducted blind to participant allocation where possible (eg, IG and CG were labeled *A* and *B* arbitrarily).

Analytical Methods

Quantitative data were analyzed descriptively using IBM SPSS 26 (IBM Corporation, released 2019). Initial analyses involved tabulated and graphical summaries of primary and secondary outcomes for each randomized group using means and variance, including CIs and SDs, and number and percentages for categorical variables to describe the full range of data at baseline and postprogram. An intention-to-treat (ITT) analysis was carried out, where missing data were rectified using the last observation carried forward (LOCF) [46]. In line with CONSORT (Consolidated Standards of Reporting Trials) guidelines [47], a per-protocol (PP) analysis was also performed on secondary outcome data from intervention completers and is reported in the *AncillaryAnalyses* section.

The study was not powered to perform inferential statistical analyses, and so to signal efficacy, we report pre- and postprogram mean differences and CIs for scores on key secondary outcome measures for the IG and CG.

Protocol

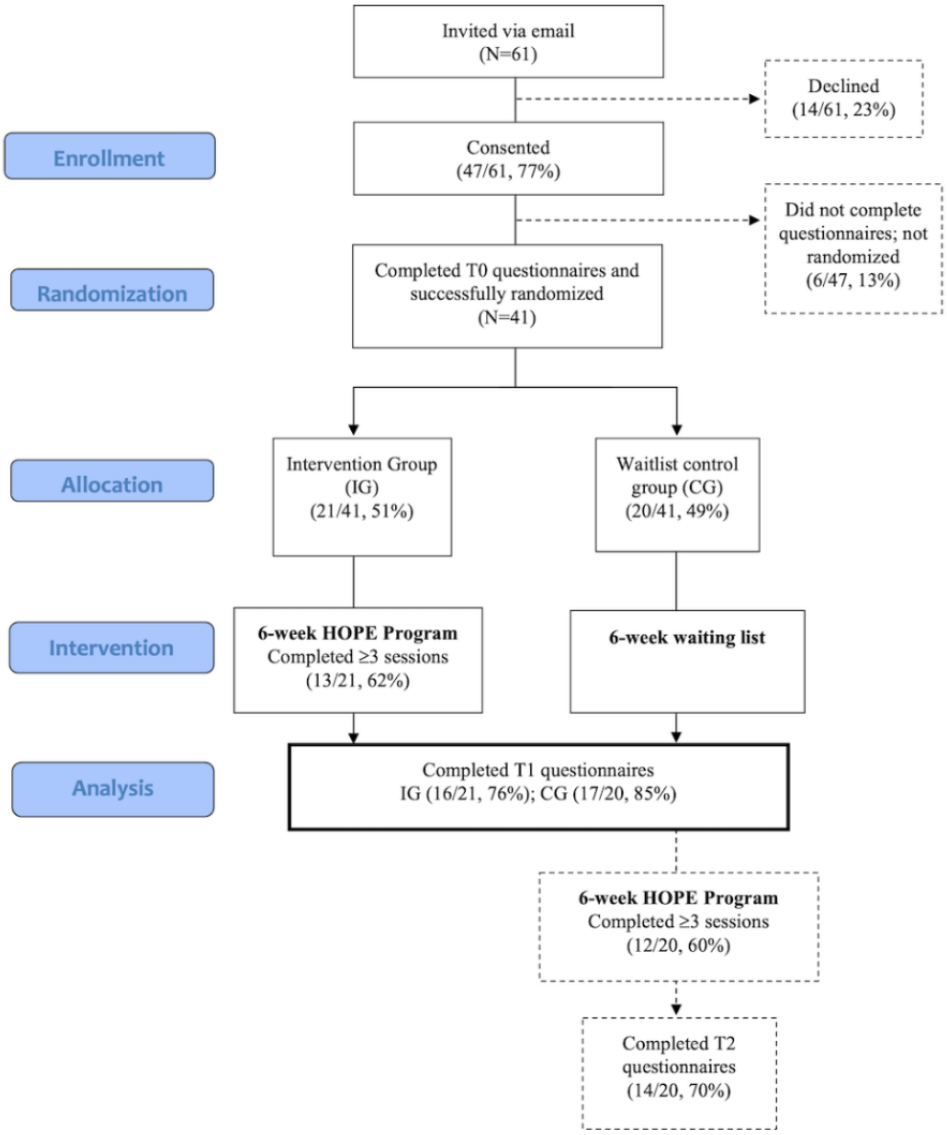
The feasibility trial protocol has been registered and published (International Registered Report Identifier IRRID: DERR1-10.2196/24264) [24].

Results

Participant Flow

Figure 1 provides the details of the participant flow through the study.

Figure 1. Participant flow through the study. HOPE: Help to Overcome Problems Effectively.



Recruitment

Recruitment started on April 30, 2020, and ended on May 5, 2020. Data collection started on April 30, 2020, for T0 baseline questionnaires and finished on September 2, 2020, for T2 follow-up questionnaires (CG only), which was 4 weeks after the end of the intervention for the CG as specified in the trial protocol.

Baseline Data

Sociodemographic and health information collected at baseline (T0) for the whole group and by treatment group are presented in Table 2. The sample consisted mostly of White participants (36/41, 87%), women (32/41, 78%) with an average age of 54.3 years. More than half of the sample (24/41, 58%) had postschool qualifications. Most participants were married or living with their partner (30/41, 73%), and more than half were employed (21/41, 51%), with just under half reporting that they had to reduce their working hours because of their cancer diagnosis (20/41, 49%). This variable was the most disproportionate across the trial arms, with more than twice as many reports of cutting work hours in the IG (14/21, 67%) than in the CG (6/20, 30%) The most commonly reported type of cancer was breast cancer (17/41, 41%), with less than half of participants still undergoing treatment (17/41, 41%). Participants may have reported more than one type of cancer; therefore, the sum of the group percentages may be >100%.

Table 2. Baseline characteristics for the whole sample and by trial arm.

Variable	All (N=41)	IG[a] (n=21)	CG[b] (n=20)
Age, mean (SD)	54.3 (11)	52.6 (11)	56.2 (12)
Female, n (%)	32 (78)	17 (81)	15 (75)
White ethnicity, n (%)	36 (88)	19 (90)	17 (85)
Married or living with partner, n (%)	30 (73)	17 (81)	13 (65)
Employed, n (%)	21 (51)	11 (52)	10 (50)
Cut work hours due to cancer, n (%)	20 (49)	14 (67)	6 (30)
Possessed postschool qualifications, n (%)	24 (58)	12 (57)	12 (60)
Still undergoing treatment for cancer, n (%)	17 (41)	9 (43)	8 (40)
Cancer type, n (%)			
Breast	17 (41)	8 (38)	9 (45)
Gynecological	4 (10)	0 (0)	4 (20)
Prostate	2 (5)	0 (0)	2 (10)
Lung	3 (7)	2 (9)	1 (5)
Colorectal	3 (7)	3 (14)	0 (0)
Gastrointestinal	2 (5)	0 (0)	2 (10)
Bladder or kidney	1 (2)	0 (0)	1 (5)
Head or neck	4 (19)	2 (9)	2 (5)
Other	9 (22)	8 (38)	1 (5)

[a]IG: intervention group.

[b]CG: control group.

Numbers Analyzed

The total number of participants enrolled in the study was 41, with 21 in the IG and 20 in the CG. All participants completed baseline (T0) questionnaires, and missing data in T1 and T2 questionnaires were populated with the LOCF method for ITT analysis. Therefore, the entire sample was included in the ITT analysis (ITT, N=41; IG, n=21; CG, n=20). The numbers for the PP analysis are detailed in the *AncillaryAnalyses* section.

Outcomes and Estimation

We describe the results for each primary and secondary outcome. Primary outcome measures pertaining to recruitment, questionnaire, and intervention completion rates are presented in Figure 1, Table 3, and Table 4. Secondary outcome measures are presented in Table 5 for the whole group (n=41) and each trial arm (IG, n=21; CG, n=20).

Table 3. Number of participants who attended none, some, or all of the Help to Overcome Problems Effectively Program sessions, for the whole group and each trial arm.

Number of sessions attended	All (N=41), n (%)	IG[a] (n=21), n (%)	CG[b] (n=20), n (%)
0	5 (12)	2 (9)	3 (15)
1	5 (12)	3 (14)	2 (10)
2	5 (12)	3 (14)	2 (10)
3	2 (5)	1 (5)	1 (5)
4	1 (2)	0 (0)	1 (5)
5	2 (5)	2 (9)	0 (0)
6	21 (51)	10 (47)	11 (55)

[a]IG: intervention group.

[b]CG: control group.

Table 4. An overview of engagement and adherence with the Help to Overcome Problems Effectively Program, for the whole group and by trial arm.

Engagement measure	All (N=41), n (%)	All (n=41), mean (SD)	IG[a] (n=21), n (%)	IG (n=21), mean (SD)	CG[b] (n=20), n (%)	CG (n=20), mean (SD)
Mean pages viewed in session 1 (range 0-16)	12.2 (76)	12.2 (6.4)	12.2 (76)	12.2 (6.2)	12.2 (76)	12.2 (6.7)
Mean pages viewed in session 2 (range 0-13)	8.5 (65)	8.5 (6.0)	8.0 (61)	8.0 (6.1)	9.1 (70)	9.1 (6.0)
Mean pages viewed in session 3 (range 0-17)	10.3 (60)	10.3 (8.0)	9.7 (57)	9.7 (8.0)	10.9 (64)	10.9 (8.2)
Mean pages viewed in session 4 (range 0-14)	7.5 (53)	7.5 (6.8)	7.0 (50)	7.0 (7.0)	8.1 (58)	8.1 (6.9)
Mean pages viewed in session 5 (range 0-16)	8.2 (51)	8.2 (8.0)	7.7 (48)	7.7 (8.0)	8.8 (55)	8.8 (8.2)
Mean pages viewed in session 6 (range 0-10)	5.1 (51)	5.1 (4.6)	5.0 (50)	5.0 (4.4)	5.3 (53)	5.3 (4.9)
Gratitude entries across whole program (range 0-9)	N/A[c]	1.5 (1.9)	N/A	1.3 (1.3)	N/A	1.7 (2.4)
Goals set across whole program (range 0-8)	N/A	2.1 (2.1)	N/A	1.8 (1.8)	N/A	2.4 (2.4)
Goal feedback given across whole program (range 0-5)	N/A	0.5 (1.1)	N/A	0.4 (1.0)	N/A	0.7 (1.2)
Likes given across whole program (range 0-56)	N/A	6.8 (12.2)	N/A	3.8 (6.6)	N/A	10.0 (15.8)
Comments posted across whole program (range 0-35)	N/A	6.8 (9.9)	N/A	5.8 (9.2)	N/A	7.9 (10.7)

[a]IG: intervention group.

[b]CG: control group.

[c]N/A: not applicable.

Table 5. All scores, intention-to-treat, on secondary outcome measures for intervention control (IG) and control group (CG), and change in scores (T1-T0), and mean difference in changes scores (IG-CG).

Secondary outcome measure	IG (n=21)			CG (n=20)			Difference in change scores Δ IG-CG, mean difference (95% CI)
	T0, mean (SD)	T1[a], mean (SD)	Postprogram change Δ (T1-T0), mean difference (95% CI)	T0, mean (SD)	T1[b], mean (SD)	Control change Δ (T1-T0), mean difference (95% CI)	
WEMWBS[c]	43.3 (9.6)	46.3 (11.7)	3.0 (−0.2 to 6.2)	43.4 (12.4)	45.1 (11.8)	1.7 (−1.6 to 5.0)	1.3 (−3.1 to 5.7)
PHQ-9[d]	10.0 (5.4)	8.1 (5.7)	−1.8 (−3.3 to −0.4)	9.2 (6.3)	9.3 (6.5)	0.1 (−2.0 to 2.2)	−1.9 (−4.4 to 0.6)
GAD-7[e]	8.8 (5.6)	7.6 (5.6)	−1.2 (−3.0 to 0.6)	9.0 (6.6)	7.4 (4.8)	−1.6 (−3.7 to 0.5)	0.4 (−2.2 to 3.1)
PAM[f]	58.8 (17.2)	60.8 (17.2)	2.0 (0.1 to 3.9)	61.3 (15.3)	59.1 (14.2)	−2.2 (−6.6 to 2.2)	4.2 (−0.3 to 8.7)

[a]LOCF: last observation carried forward, n=5.

[b]LOCF: last observation carried forward, n=3.

[c]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]GAD-7: 7-item Generalized Anxiety Disorder scale.

[f]PAM: Patient Activation Measure.

Primary Outcomes

Recruitment rates

The recruitment rate for this feasibility study was 77% (47/61). MCS referred to 61 participants for the study, and 77% of participants (47/61) provided digital informed consent. Only participants who (1) consented, (2) completed baseline questionnaires, and (3) were randomized, were included in this feasibility study (n=41). Six participants provided informed consent but did not complete the T0 questionnaire; therefore, they were not randomized (6/47, 13%). The study participants who completed the baseline questionnaires (n=41) were randomized to either the IG (n=21) or the CG (n=20) groups.

Retention, Follow-up, and Completion Rates

The retention rate across the sample was 51%, with more than half of the participants attending all 6 program sessions (whole group: 21/41, 51%; IG: 10/21, 47%; CG: 11/20, 55%).

Across the whole group, there was a follow-up rate of 80% (33/41 participants completed T1 questionnaires). Across the trial arms, the follow-up rate was 76% (16/21) in the IG and 85% (17/20) in the CG.

The completion rates (3 sessions and T1 questionnaire) for the whole group were 61% (25/41), 62% (13/21) in the IG, and 60% (12/20) in the CG.

Adherence and engagement measures

Table 4 shows the selection of the engagement data collected by the intervention platform. The mean number of pages viewed per session generally decreased as the program progressed and ranged from 12.2/16 (76%) in session 1 to 5.1/10 (51%) in session 6, across the whole group. The mean pages viewed in each session were consistent across the entire group and both trial arms for sessions 1 and 6. The mean number of pages viewed was slightly higher for the CG than for sessions 2 to 5. Furthermore, the CG tended to set slightly more goals, give more likes, and post more comments than the IG, on average, across the course of intervention. There was a negligible difference in the mean gratitude entries and goal feedback between the 2 trial arms.

Sample Size and Effect Size Estimation for Future Definitive Trial

To guide the sample size estimation for a future definitive trial, we used the results of this study. Accordingly, we calculated the expected minimum effect sizes using the primary outcome variable, WEMWBS, and based on the present data, considering the mean change scores for IG and CG and a moderate between-group effect size. Given an *a priori* α of .05, 87 participants were required per group to minimally detect moderate effect sizes (ie, Cohen $f \geq 0.25$; Cohen $d \geq 0.5$), or 42 participants per group to minimally detect large effect sizes (Cohen $f \geq 0.4$; Cohen $d \geq 0.8$), both with a power of 0.95 [36]. Indeed, we must caveat this guide with acknowledgment that the inherent imprecision in between treatment group effect size estimates from studies with small samples can be high and thus must be considered pragmatically in future trials [48].

Progression Criteria

All predetermined progression criteria for a definitive trial were met, as described in the sections above.

Secondary Outcomes

An ITT analysis was carried out, with missing data rectified by LOCF. The mean scores and SDs for the secondary outcome measures for both trial arms are presented in Table 5. On average, participants in the IG made small improvements from T0 to T1. Participants in the CG showed little to no improvement during the same period. However, after attending the HOPE program (ie, T2), the CG made greater improvements in all secondary measures than the IG.

Table 6 shows the number and proportion of participants exceeding the cutoff scores for probable clinical depression and

anxiety at each time point across the trial. In the IG, the number of participants reporting probable clinical levels of depression decreased by 1 from preprogram (8/21, 38%) to postprogram (7/21, 33%), and cases of depression remained the same pre- and postprogram (10/21, 47%). In the CG, at the point of entry into the trial (T0), the proportion of participants reporting probable clinical levels of depression and anxiety was the same (10/20, 50%). From pre- to postprogram in the CG, there was a 15% decrease in cases of probable depression (T1, 7/20, 35%; T2, 4/20, 20%) and anxiety (T1, 8/20, 40%; T2, 5/20, 25%).

Table 6. Proportion of participants reporting probable clinical levels of depression and anxiety at each time point across the trial.

Secondary outcome measure	IG[a], n (%); n=21		CG[b], n (%); n=20		
	T0	T1[c]	T0	T1[d]	T2[e]
Cases of probable depression	8 (38)	7 (33)	10 (50)	7 (35)	4 (20)
Cases of probable clinical anxiety	10 (47)	10 (47)	10 (50)	8 (40)	5 (25)

[a]IG: intervention group.

[b]CG: control group.

[c]LOCF: last observation carried forward, n=5.

[d]LOCF: last observation carried forward, n=3.

[e]LOCF: last observation carried forward, n=6.

Ancillary Analyses

We conducted a PP analysis, which included only those participants who completed all study questionnaires and attended at least 3 intervention sessions (PP whole sample n=25; IG n=13, CG n=12). Table 7 shows the secondary outcome measures for the participants. The data in Table 7 show patterns similar to those of the ITT in Table 5. On average, participants in the IG showed modest improvements from T0 to T1. Participants in the CG showed little to no improvement during the same period. The PP analysis shows a difference in the change scores (final column) of greater magnitude compared with the ITT analysis.

Table 7. Scores on secondary outcome measures for intervention group (IG; n=13) and control group (CG; n=12) intervention completers (PP), including mean difference in change scores (IG-CG).

Secondary outcome measure	IG (n=13)			CG (n=12)			Difference in change scores Δ IG-CG, mean difference (95% CI)
	T0, mean (SD)	T1, mean (SD)	Postprogram change Δ (T1-T0), mean difference (95% CI)	T0, mean (SD)	T1, mean (SD)	Control change Δ (T1-T0), mean difference (95% CI)	
WEMWBS[a]	44.9 (10.1)	49.5 (12.7)	4.6 (−0.5 to 9.7)	43.8 (12.0)	42.7 (11.7)	−1.1 (−4.7 to 2.5)	5.7 (−0.3 to 11.7)
PHQ-9[b]	9.1 (6.1)	6.8 (6.5)	−2.2 (−4.3 to −0.2)	9.1 (6.6)	10.8 (6.7)	1.7 (0.4 to 2.9)	−3.9 (−6.2 to −1.6)
GAD-7[c]	8.6 (6.4)	7.0 (6.6)	−1.6 (−4.4 to 1.2)	8.5 (5.7)	8.8 (4.8)	0.3 (−1.8 to 2.3)	−1.9 (−5.2 to 1.5)
PAM[d]	63.2 (15.2)	66.2 (14.6)	3.0 (0 to 6.0)	61.6 (12.8)	56.1 (11.2)	−5.5 (−11.7 to 0.7)	8.5 (2.2 to 14.8)

[a]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]GAD-7: 7-item Generalized Anxiety Disorder scale.

[d]PAM: Patient Activation Measure.

Harms

In line with the trial protocol [24], participants who indicated self-harm or suicidal thoughts on the PHQ-9 measure were contacted, along with the MCS administrator, by Hope for the Community (H4C) and were provided with the contact details of local mental health agencies and Samaritans and encouraged to visit their general practitioner. This was the case for 22% (9/41) of participants' preprogram, and 9% (4/41) of participants postprogram (data not shown in tables). At postprogram, there were no participants who indicated self-harm or suicidal thoughts where they had not already indicated this at the preprogram.

As detailed in the Methods section, participants scoring 10 on the PHQ-9 or 8 on the GAD-7 were categorized as having

reached a probable clinical level of depression or anxiety, respectively. Depression was indicated in 44% (18/41) of participants at preprogram and 34% (14/41) at T1. Anxiety was indicated in 49% (20/41) of participants at preprogram and 44% (18/41) at T1. In line with the trial protocol [24], all of these participants were contacted by H4C and encouraged to visit their general practitioner and were signposted to further sources of support as listed above.

At postprogram, there were no participants who reported a probable clinical level of depression where they had not already reported this at the preprogram. However, at postprogram, 5% (2/41) of participants reached a probable clinical level of anxiety but were not previously at this level in the preprogram. Both participants scored 7 on the GAD-7 measure at preprogram, increasing to scores of 8 (n=1) and 9 (n=1) at postprogram. Both participants were contacted by H4C, as outlined above, and in the trial protocol [24]. To provide further context, both participants were in the IG, and only attended 1 (n=1) or 2 (n=1) sessions of the intervention. Both participants were still undergoing treatment for cancer, and one described significant personal stress unrelated to their cancer. Although we cannot rule out the possibility that the intervention may have caused increased anxiety in these 2 participants, they did not engage in more than 2 sessions of the intervention, and the context of the COVID-19 pandemic is linked to increased anxiety among patients with cancer [1,3,5,6]. Furthermore, other participants showed positive changes in their pre- and postprogram mental well-being scores.

Discussion

Principal Findings

The feasibility RCT of the digital HOPE program aimed to assess primary outcomes measuring trial feasibility and secondary outcomes relating to measures of participant well-being. The trial yielded encouraging data on the primary outcome measures of recruitment, retention, follow-up, adherence, and engagement rates. More than three-quarters of the participants invited (47/61, 77%) were willing to provide consent and be randomized to either the HOPE program starting the following week, or to a 6-week waiting list. Just over half of the sample (ie, IG and CG combined; 21/41, 51%) completed all 6 sessions of the intervention, and almost two-thirds of the sample (26/41, 63%) completed at least 3 sessions (note that n=1 did not complete the T1 questionnaire, so it was not categorized as an intervention completer. The follow-up rate was encouraging, with a large proportion of participants completing the study questionnaires at T1 (33/41, 80%). Of the participants who completed the T1 questionnaires, 25 also attended 3 intervention sessions, meeting the criteria for intervention completion (25/41, 61%). In terms of engagement, within the sessions, participants viewed between half and three-quarters of the content, on average (range 76%-51%). All of the predetermined progression criteria were met, confirming that a full-scale, fully powered RCT of the digital HOPE program for people with cancer would be feasible.

On average, participants showed increased scores on positive mental well-being and patient activation and decreased scores on anxiety and depression at postprogram relative to baseline. We did not ask participants specific questions relating to their well-being during the COVID-19 pandemic. However, we can tentatively compare data from this study with a previous cohort of people with cancer in the digital HOPE program, collected before the COVID-19 pandemic [23]. The proportion of participants reporting probable clinical levels of depression at baseline was slightly higher in the prepandemic cohort [23] than in the current trial (24/51, 47%, and 18/41, 44%, respectively). However, at baseline, probable clinical levels of anxiety were lower in the prepandemic cohort (22/51, 43%) than in the current trial (20/41, 49%). Indeed, in the current trial, baseline GAD-7 scores for anxiety 8.9 (SD 6.0; pooled data not shown in *Results*) were higher than those in our prepandemic study 6.8 (SD 4.9) [23]. Given the widespread reports of the negative effect of the COVID-19 pandemic on the mental health of people with cancer [3], the baseline scores in the current trial may represent elevated COVID-19–related anxiety.

Exploration of the secondary outcome data highlights the safety of the HOPE program as a digital self-management intervention for people with cancer. There were no cases of participants reporting increased symptoms of depression, and only 2 participants reported increased postprogram anxiety. There were no cases where participants reported thoughts of suicide or harming themselves at postprogram, where they had not already reported this at baseline.

Limitations

This study found that overall engagement, measured by the percentage of pages viewed, seemed to decline as participants progressed through the sessions. This may be due to fatigue or redundant content. Qualitative investigation into what content participants engaged with, and elements they found more or less relevant or helpful would be a useful supplement to improve the intervention.

This feasibility RCT was not powered to detect statistically significant differences in pre- and postscores on secondary outcomes. However, the results indicate that the HOPE program has the potential to have a positive effect on mental well-being, depression, and anxiety in people with cancer. These have been identified as important outcomes for people with cancer [9-11] and echo the results of a previous pre- and poststudy of the HOPE program [23], giving further confidence in the potential efficacy of the intervention. Data from a fully powered trial will allow us, for the first time, to report statistically significant differences in pre- and postprogram scores for both IG and CG. However, unless we account for expectancy effects, we cannot be sure about the efficacy of the intervention [49]. Therefore, future trials will need to use an appropriate active control program, which equates expectations to those of the IG, to allow a causal conclusion about the effectiveness of the intervention effectiveness. There are a number of ways to address this issue in future work. First, empirical measures of expectancy could be collected before and after the program to assess for any correlation between expectancy and improvement in either group. Although this would not reduce any expectancy effects, it would allow us to account for these in statistical analyses, interpretation of the data, and evaluation of the intervention.

Second, in a future trial, we could use an active control program, such as an alternative digital self-management intervention, or a modified version of the HOPE program (eg, without the goal-setting feature; self-directed course). Active control needs to be carefully matched to the intervention in terms of expectancy, content, and interaction [49]. Researchers suggest that if participants' well-being is improved postintervention, then the mechanisms by which this benefit occurs are irrelevant. However, it becomes a problem if the improvement is only detectable in the laboratory, or in this case, at the time of the postprogram data collection. A third way to address expectancy in future trials would be to implement further follow-up time points (eg, 3, 6, and 12 months postrandomization) to assess whether improvements to participant well-being persist in the longer term. Ideally, in future trials, we would use all 3 methods to reduce placebo effects and increase confidence in the treatment effect of the HOPE program.

Generalizability

The recruitment for this feasibility RCT was from an opportunity sample of self-selecting participants referred by MCS. The self-selecting nature of the recruitment strategy may yield participants who are generally more motivated to seek help and/or help themselves. However, this recruitment strategy facilitated the rapid attainment of trial recruitment targets in this study [50]. Research has shown that recruitment via social media is more effective if advertised by a collaborative cancer charity [51]. In this respect, in the current climate of increased need for digital research and provision of self-management support, we optimized our recruitment strategy in this feasibility trial and will adopt this again in the definitive trial.

Most participants were White (36/41, 88%), female (32/41, 78.0%), married (30/41, 73%), and educated (24/41, 58%), and the most commonly reported type of cancer was breast cancer (17/41, 41%). This likely relates to the demographics of people who engage with the MCS charity. Although this may limit the generalizability of the results to other demographic groups, some aspects are in line with wider population statistics and research findings. The 2011 Census [52] reported that 86% of the population in England and Wales was White, and therefore, the sample in this study is representative of the wider population in this respect. Breast cancer is the most common type of cancer in the United Kingdom, accounting for 15.1% of malignant cancer registrations in England in 2017 [53]; however, 41.5% of participants in this study reported breast cancer. The data presented in this study may not be representative of other cancer populations. As such, the efficacy signal and feasibility findings of this study may not be generalizable to other types of cancer, or to non-White men, for example. We will seek advice from our partners and trial experts before proceeding to a definitive trial. It may be more appropriate to run a definitive RCT of the HOPE program for breast cancer survivors only, as (1) breast cancer is the most commonly diagnosed cancer in the United Kingdom, and (2) our own data, for example, 23 and unpublished studies, show that it is mainly women with breast cancer who participate in the HOPE program. However, the HOPE program was designed to help people living with all types of cancer; thus, the community HOPE program run by MCS will continue to be open to all cancer survivors.

A low attendance rate for men is common in self-management and is linked to their reluctance to seek help [54]. In terms of recruitment, men are more likely to respond to marketing and recruitment messages that emphasize stoicism, independence, and control [54] and where the materials contain images of men [55]. Once recruited, there are also qualitative differences in how men and women engage with their peers in the same- or mixed-sex web-based cancer support groups [56]. A recent systematic review confirmed that men are more oriented toward informational support and women toward emotional support [57]. In terms of the current intervention, further intervention development is required to ensure the relevance and acceptability of the intervention and potentially to co-design tailored versions for more diverse groups and communities. This may require (1) further consultation with MCS to co-design specific programs for gendered cancers (eg, a HOPE program for testicular, prostrate, or breast cancer); (2) co-development of course content and recruitment materials to increase the engagement of men in a general cancer intervention; and (3) partnering with different charities to enhance engagement with people with cancer from different ethnic groups, socioeconomic status, and educational attainment. A future trial could examine the feasibility of recruitment through the NHS, from clinics, consultation rooms, or waiting rooms, to broaden the recruitment strategy to wider communities. For future cohorts, we will encourage MCS to review their recruitment materials to ensure that they contain images and messages that appeal to multiple audiences and are advertised in (largely web-based) areas and locations frequented by people of all ages, ethnicities, genders, and income groups [54-57].

Interpretation

The digital HOPE program is a feasible self-management intervention for people with cancer, although almost half of our sample comprised people with breast cancer. All progression criteria were met, providing support for a full-scale definitive trial. However, caution must be taken when interpreting the generalizability of our feasibility estimates, and a further discussion with our research trial group will be undertaken to determine the appropriate action for progression. Generally, at postprogram, all participants showed increased scores on positive mental well-being and patient activation and decreased scores on anxiety and depression relative to baseline, signaling intervention efficacy. Minimal harm was indicated, and no participants reported postprogram symptoms of anxiety or depression that were not present at baseline, and similarly for thoughts of suicide and self-harm.

The asynchronous nature of the course and the autonomy afforded to participants means that they can broadly tailor the course to accommodate personal needs. Participants can attend sessions that are interesting or relevant to them, essentially creating their own person-centered support by self-selecting the content. Rather than interpreting attrition as a negative trial outcome, we concur with the view that participants can be *eAttainers* [35,58]. That is, participants take what they need from the course and then may appear to *drop out* but are nonetheless satisfied and fulfilled by the content they received. Participants do not need to engage with all available components of the intervention, or all use the same components, as goals

differ across individuals [35]. For example, some participants obtain reassurance from sharing their challenges and concerns with others, a common group curative factor described as "universality" by Yalom [59]. This approach aligns with the tailored nature of the HOPE program, enabling users to self-select the most relevant intervention tools, features, or content for themselves. This is supported by the overwhelmingly positive user evaluations collected postprogram in this feasibility trial (data not shown) and in a previous HOPE program study [23]. As the focus of this paper was to assess the feasibility of an RCT, and not to evaluate the efficacy and acceptability of the HOPE program [48], we did not present participant feedback data here.

Conclusions

The advent of the digital HOPE program has coincided with requests from academics, leading cancer charities, and the NHS, for more person-centered cancer self-management, especially during the recent global pandemic [2,3,8]. As such, there has been rapid and essential growth in the provision of health care digitally to allow remote care [60,61], and these advances in digital health care will still be viable in the postpandemic era. This feasibility RCT suggests that the digital HOPE program could supplement in-person emotional and psychological support for people with cancer, offering greater choice and flexibility in accessing support. Although digital cancer self-management interventions, such as the HOPE program, have boomed during COVID-19 [60-62], more data are needed to suggest whether digital could be considered an alternative, as well as complementary, format to in-person intervention programs. As the HOPE program content can be delivered on the web or in-person, it offers flexibility and a choice of formats for participants when social distancing measures are eased after the pandemic.

Acknowledgments

The research study was investigator funded and approved by the Coventry University Ethics Committee (P106024) on April 28, 2020.

Conflicts of Interest

AT is the co-inventor of the HOPE Program. GM is the CEO of Hope For The Community (H4C).

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 513 KB - jmir_v23i11e28322_app1.pdf]

References

1. Lai AG, Pasea L, Banerjee A, Hall G, Denaxas S, Chang WH, et al. Estimated impact of the COVID-19 pandemic on cancer services and excess 1-year mortality in people with cancer and multimorbidity: near real-time data on cancer care, cancer deaths and a population-based cohort study. BMJ Open 2020 Nov 17;10(11):e043828 [FREE Full text] [doi: 10.1136/bmjopen-2020-043828] [Medline: 33203640]
2. Raymond E, Thieblemont C, Alran S, Faivre S. Impact of the COVID-19 outbreak on the management of patients with cancer. Target Oncol 2020 Jun;15(3):249-259 [FREE Full text] [doi: 10.1007/s11523-020-00721-1] [Medline: 32445083]
3. Lost in lockdown. Macmillan Cancer Support. URL: https://medium.com/macmillan-press-releases-and-statements/lost-in-lockdown-5cefcbb8d76b [accessed 2021-02-17]
4. Burki TK. Cancer care in the time of COVID-19. Lancet Oncol 2020 May;21(5):628 [FREE Full text] [doi: 10.1016/S1470-2045(20)30201-1] [Medline: 32213339]
5. Nekhlyudov L, Duijts S, Hudson SV, Jones JM, Keogh J, Love B, et al. Addressing the needs of cancer survivors during the COVID-19 pandemic. J Cancer Surviv 2020 Oct;14(5):601-606 [FREE Full text] [doi: 10.1007/s11764-020-00884-w] [Medline: 32335850]
6. Richards M, Anderson M, Carter P, Ebert BL, Mossialos E. The impact of the COVID-19 pandemic on cancer care. Nat Cancer 2020 May 20:1-3 (forthcoming) [FREE Full text] [doi: 10.1038/s43018-020-0074-y] [Medline: 32838302]
7. Chan A, Ashbury F, Fitch MI, Koczwara B, Chan RJ, MASCC Survivorship Study Group. Cancer survivorship care during COVID-19-perspectives and recommendations from the MASCC survivorship study group. Support Care Cancer 2020 Aug;28(8):3485-3488 [FREE Full text] [doi: 10.1007/s00520-020-05544-4] [Medline: 32451702]
8. The NHS long term plan. National Health Service. URL: https://www.longtermplan.nhs.uk/ [accessed 2021-02-17]
9. Armes J, Crowe M, Colbourne L, Morgan H, Murrells T, Oakley C, et al. Patients' supportive care needs beyond the end of cancer treatment: a prospective, longitudinal survey. J Clin Oncol 2009 Dec 20;27(36):6172-6179. [doi: 10.1200/JCO.2009.22.5151] [Medline: 19884548]
10. Foster C, Fenlon D. Recovery and self-management support following primary cancer treatment. Br J Cancer 2011 Nov 08;105 Suppl 1:21-28 [FREE Full text] [doi: 10.1038/bjc.2011.419] [Medline: 22048029]

11. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. Br J Cancer 2011 Nov 08;105 Suppl 1:11-20 [FREE Full text] [doi: 10.1038/bjc.2011.418] [Medline: 22048028]
12. Fabi A, Bhargava R, Fatigoni S, Guglielmo M, Horneber M, Roila F, ESMO Guidelines Committee. Electronic Address: clinicalguidelines@esmo.org. Cancer-related fatigue: ESMO Clinical Practice Guidelines for diagnosis and treatment. Ann Oncol 2020 Jun;31(6):713-723 [FREE Full text] [doi: 10.1016/j.annonc.2020.02.016] [Medline: 32173483]
13. Hofman M, Ryan JL, Figueroa-Moseley CD, Jean-Pierre P, Morrow GR. Cancer-related fatigue: the scale of the problem. Oncologist 2007;12 Suppl 1:4-10 [FREE Full text] [doi: 10.1634/theoncologist.12-S1-4] [Medline: 17573451]
14. Simard S, Thewes B, Humphris G, Dixon M, Hayden C, Mireskandari S, et al. Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies. J Cancer Surviv 2013 Sep;7(3):300-322. [doi: 10.1007/s11764-013-0272-z] [Medline: 23475398]
15. Simard S, Savard J, Ivers H. Fear of cancer recurrence: specific profiles and nature of intrusive thoughts. J Cancer Surviv 2010 Dec;4(4):361-371. [doi: 10.1007/s11764-010-0136-8] [Medline: 20617394]
16. Koch L, Bertram H, Eberle A, Holleczek B, Schmid-Höpfner S, Waldmann A, et al. Fear of recurrence in long-term breast cancer survivors-still an issue. Results on prevalence, determinants, and the association with quality of life and depression from the cancer survivorship--a multi-regional population-based study. Psychooncology 2014 May;23(5):547-554. [doi: 10.1002/pon.3452] [Medline: 24293081]
17. Koch L, Jansen L, Brenner H, Arndt V. Fear of recurrence and disease progression in long-term (≥ 5 years) cancer survivors--a systematic review of quantitative studies. Psychooncology 2013 Jan;22(1):1-11. [doi: 10.1002/pon.3022] [Medline: 22232030]
18. Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. Lancet Oncol 2013 Jul;14(8):721-732. [doi: 10.1016/S1470-2045(13)70244-4] [Medline: 23759376]
19. Inhestern L, Beierlein V, Bultmann JC, Möller B, Romer G, Koch U, et al. Anxiety and depression in working-age cancer survivors: a register-based study. BMC Cancer 2017 May 19;17(1):347 [FREE Full text] [doi: 10.1186/s12885-017-3347-9] [Medline: 28526007]
20. Niedzwiedz CL, Knifton L, Robb KA, Katikireddi SV, Smith DJ. Depression and anxiety among people living with and beyond cancer: a growing clinical and research priority. BMC Cancer 2019 Oct 11;19(1):943 [FREE Full text] [doi: 10.1186/s12885-019-6181-4] [Medline: 31604468]
21. Turner A. HOPE: a positive psychological group self-management support programme for cancer survivors. Psycho-oncology 2012;21(S2):16. [doi: 10.1111/j.1099-1611.2012.03051.x]
22. Turner A, Batehup L, Surendranath S. Hope is key to recovery. Nurs Stand 2011 Jan;25(18):20-21. [doi: 10.7748/ns.25.18.20.s26] [Medline: 21309317]
23. Martin F, Wright H, Moody L, Whiteman B, McGillion M, Clyne W, et al. Help to overcome problems effectively for cancer survivors: development and evaluation of a digital self-management program. J Med Internet Res 2020 May 19;22(5):e17824 [FREE Full text] [doi: 10.2196/17824] [Medline: 32209529]
24. Wright H, Martin F, Clyne W, Clark CC, McGillion M, Matouskova G, et al. A digital program (Hope) for people living with cancer during the COVID-19 pandemic: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2020 Dec 04;9(12):e24264 [FREE Full text] [doi: 10.2196/24264] [Medline: 33237877]
25. Seligman M. Learned Optimism: How to Change Your Mind and Your Life. UK: Hachette; 2018.
26. Seligman M. Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. New York: Atria Books; 2004.
27. Linley PA, Joseph S. Positive Psychology in Practice. Hoboken, New Jersey: Wiley; 2004.
28. Tugade MM, Fredrickson BL, Barrett LF. Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. J Pers 2004 Dec;72(6):1161-1190 [FREE Full text] [doi: 10.1111/j.1467-6494.2004.00294.x] [Medline: 15509280]
29. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004.
30. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016 Oct 21;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
31. Qualtrics. URL: https://www.qualtrics.com/au/?rid=ip&prevsite=en&newsite=au&geo=IN&geomatch=au [accessed 2021-10-01]
32. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
33. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]

34. 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]
35. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
36. Cohen J. Statistical Power Analysis for the Behavioural Sciences, Second Edition. New York: Lawrence Erlbaum Associates; 1988.
37. Hallingberg B, Turley R, Segrott J, Wight DR, Craig PE, Moore LJ, et al. Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. Pilot Feasibility Stud 2018 May 28;4:104 [FREE Full text] [doi: 10.1186/s40814-018-0290-8] [Medline: 29854417]
38. Young HM, Goodliffe S, Madhani M, Phelps K, Regen E, Locke A, et al. Co-producing progression criteria for feasibility studies: a partnership between patient contributors, clinicians and researchers. Int J Environ Res Public Health 2019 Oct 06;16(19):3756 [FREE Full text] [doi: 10.3390/ijerph16193756] [Medline: 31590424]
39. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
40. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
41. Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): group and individual level analysis. Health Qual Life Outcomes 2012 Dec 27;10:156 [FREE Full text] [doi: 10.1186/1477-7525-10-156] [Medline: 23270465]
42. 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]
43. 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]
44. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
45. Darlow S, Wen K. Development testing of mobile health interventions for cancer patient self-management: a review. Health Informatics J 2016 Sep;22(3):633-650 [FREE Full text] [doi: 10.1177/1460458215577994] [Medline: 25916831]
46. Dainty AD, Fox M, Lewis N, Hunt M, Holtham E, Timmons S, et al. A mixed methods feasibility study to evaluate the use of a low-intensity, nurse-delivered cognitive behavioural therapy for the treatment of irritable bowel syndrome. BMJ Open 2014 Jun 17;4(6):e005262 [FREE Full text] [doi: 10.1136/bmjopen-2014-005262] [Medline: 24939813]
47. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3):e1000251 [FREE Full text] [doi: 10.1371/journal.pmed.1000251] [Medline: 20352064]
48. Leon A, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
49. Boot WR, Simons DJ, Stothart C, Stutts C. The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. Perspect Psychol Sci 2013 Jul;8(4):445-454. [doi: 10.1177/1745691613491271] [Medline: 26173122]
50. Thewes B, Rietjens JA, van den Berg SW, Compen FR, Abrahams H, Poort H, et al. One way or another: the opportunities and pitfalls of self-referral and consecutive sampling as recruitment strategies for psycho-oncology intervention trials. Psychooncology 2018 Aug;27(8):2056-2059 [FREE Full text] [doi: 10.1002/pon.4780] [Medline: 29808508]
51. Hulbert-Williams NJ, Pendrous R, Hulbert-Williams L, Swash B. Recruiting cancer survivors into research studies using online methods: a secondary analysis from an international cancer survivorship cohort study. Ecancermedicalscience 2019 Dec 12;13:990 [FREE Full text] [doi: 10.3332/ecancer.2019.990] [Medline: 32010214]
52. Population of England and Wales. GOV.UK : Ethnicity Facts and Figures. 2018. URL: https://tinyurl.com/3xdc7ayr [accessed 2021-02-17]
53. Cancer registration statistics, England: 2017. Office for National Statistics. 2017. URL: https://tinyurl.com/2u353s99 [accessed 2021-02-17]
54. Galdas P, Fell J, Bower P, Kidd L, Blickem C, McPherson K, et al. The effectiveness of self-management support interventions for men with long-term conditions: a systematic review and meta-analysis. BMJ Open 2015 Mar 20;5(3):e006620 [FREE Full text] [doi: 10.1136/bmjopen-2014-006620] [Medline: 25795688]
55. Horrell LN, Kneipp SM, Ahn S, Towne SD, Mingo CA, Ory MG, et al. Chronic disease self-management education courses: utilization by low-income, middle-aged participants. Int J Equity Health 2017 Jun 27;16(1):114 [FREE Full text] [doi: 10.1186/s12939-017-0604-0] [Medline: 28655319]

56. Klemm P, Hurst M, Dearholt SL, Trone SR. Gender differences on internet cancer support groups. Comput Nurs 1999;17(2):65-72. [Medline: 10194883]
57. van Eenbergen MC, van de Poll-Franse LV, Krahmer E, Verberne S, Mols F. Analysis of content shared in online cancer communities: systematic review. JMIR Cancer 2018 Apr 03;4(1):e6 [FREE Full text] [doi: 10.2196/cancer.7926] [Medline: 29615384]
58. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
59. Yalom I, Leszcz M. Theory and Practice of Group Psychotherapy. 5th edition. New York: Basic Books; 2008.
60. Salako O, Okunade K, Allsop M, Habeebu M, Toye M, Oluyede G, et al. Upheaval in cancer care during the COVID-19 outbreak. Ecancermedicalscience 2020 Apr 1;14:ed97 [FREE Full text] [doi: 10.3332/ecancer.2020.ed97] [Medline: 32269597]
61. Lonergan PE, Washington III SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19: cross-sectional analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
62. Morrow-Howell N, Galucia N, Swinford E. Recovering from the COVID-19 pandemic: a focus on older adults. J Aging Soc Policy 2020;32(4-5):526-535. [doi: 10.1080/08959420.2020.1759758] [Medline: 32336225]

Abbreviations

CG: control group
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: 7-item Generalized Anxiety Disorder scale
H4C: Hope for the Community
HOPE: Help to Overcome Problems Effectively
IG: intervention group
ITT: intention-to-treat
LOCF: last observation carried forward
MCS: Macmillan Cancer Support
NHS: National Health Service
PHQ-9: 9-item Patient Health Questionnaire
PP: per-protocol
RCT: randomized controlled trial
WEMWBS: Warwick Edinburgh Mental Well-being Scale

Edited by G Eysenbach; submitted 01.03.21; peer-reviewed by J Farragher, L Horrell; comments to author 31.03.21; revised version received 25.05.21; accepted 02.08.21; published 05.11.21.

Please cite as:
Wright H, Martin F, Clyne W, Clark CCT, Matouskova G, McGillion M, Turner A
A Digital Self-management Program (Help to Overcome Problems Effectively) for People Living With Cancer: Feasibility Randomized Controlled Trial
J Med Internet Res 2021;23(11):e28322
URL: https://www.jmir.org/2021/11/e28322
doi:10.2196/28322
PMID:34738912

Original Paper

A Smartphone Intervention for People With Serious Mental Illness: Fully Remote Randomized Controlled Trial of CORE

Dror Ben-Zeev[1], PhD; Ayesha Chander[1], MRes; Justin Tauscher[1], PhD; Benjamin Buck[1], PhD; Subigya Nepal[2], BSc; Andrew Campbell[2], PhD; Guy Doron[3], PhD

[1]Behavioral Research in Technology and Engineering Center, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[2]Department of Computer Science, Dartmouth College, Hanover, NH, United States

[3]School of Psychology, Interdisciplinary Center, Herzliya, Israel

Corresponding Author:
Dror Ben-Zeev, PhD
Behavioral Research in Technology and Engineering Center
Department of Psychiatry and Behavioral Sciences
University of Washington
1959 NE Pacific Street
Seattle, WA, 98195
United States
Phone: 1 206 685 9655
Email: dbenzeev@uw.edu

Abstract

Background: People with serious mental illness (SMI) have significant unmet mental health needs. Development and testing of digital interventions that can alleviate the suffering of people with SMI is a public health priority.

Objective: The aim of this study is to conduct a fully remote randomized waitlist-controlled trial of CORE, a smartphone intervention that comprises daily exercises designed to promote reassessment of dysfunctional beliefs in multiple domains.

Methods: Individuals were recruited via the web using Google and Facebook advertisements. Enrolled participants were randomized into either active intervention or waitlist control groups. Participants completed the Beck Depression Inventory-Second Edition (BDI-II), Generalized Anxiety Disorder-7 (GAD-7), Hamilton Program for Schizophrenia Voices, Green Paranoid Thought Scale, Recovery Assessment Scale (RAS), Rosenberg Self-Esteem Scale (RSES), Friendship Scale, and Sheehan Disability Scale (SDS) at baseline (T1), 30-day (T2), and 60-day (T3) assessment points. Participants in the active group used CORE from T1 to T2, and participants in the waitlist group used CORE from T2 to T3. Both groups completed usability and accessibility measures after they concluded their intervention periods.

Results: Overall, 315 individuals from 45 states participated in this study. The sample comprised individuals with self-reported bipolar disorder (111/315, 35.2%), major depressive disorder (136/315, 43.2%), and schizophrenia or schizoaffective disorder (68/315, 21.6%) who displayed moderate to severe symptoms and disability levels at baseline. Participants rated CORE as highly usable and acceptable. Intent-to-treat analyses showed significant treatment×time interactions for the BDI-II ($F_{1,313}$=13.38; P<.001), GAD-7 ($F_{1,313}$=5.87; P=.01), RAS ($F_{1,313}$=23.42; P<.001), RSES ($F_{1,313}$=19.28; P<.001), and SDS ($F_{1,313}$=10.73; P=.001). Large effects were observed for the BDI-II (d=0.58), RAS (d=0.61), and RSES (d=0.64); a moderate effect size was observed for the SDS (d=0.44), and a small effect size was observed for the GAD-7 (d=0.20). Similar changes in outcome measures were later observed in the waitlist control group participants following crossover after they received CORE (T2 to T3). Approximately 41.5% (64/154) of participants in the active group and 60.2% (97/161) of participants in the waitlist group were retained at T2, and 33.1% (51/154) of participants in the active group and 40.3% (65/161) of participants in the waitlist group were retained at T3.

Conclusions: We successfully recruited, screened, randomized, treated, and assessed a geographically dispersed sample of participants with SMI entirely via the web, demonstrating that fully remote clinical trials are feasible in this population; however, study retention remains challenging. CORE showed promise as a usable, acceptable, and effective tool for reducing the severity of psychiatric symptoms and disability while improving recovery and self-esteem. Rapid adoption and real-world dissemination

of evidence-based mobile health interventions such as CORE are needed if we are to shorten the science-to-service gap and address the significant unmet mental health needs of people with SMI during the COVID-19 pandemic and beyond.

Trial Registration: ClinicalTrials.gov NCT04068467; https://clinicaltrials.gov/ct2/show/NCT04068467

(J Med Internet Res 2021;23(11):e29201) doi:10.2196/29201

KEYWORDS

mobile health; schizophrenia; bipolar disorder; depression; mobile phone

Introduction

Background

People with serious mental illnesses (SMIs), including schizophrenia-spectrum disorders, bipolar disorder, and severe and persistent depression, experience significant psychiatric symptoms such as hallucinations, delusions, and severe mood episodes. SMI is often accompanied by functional and psychosocial impairments, housing and employment challenges, and poverty [1-3]. Treatment of people with SMI typically takes place in publicly funded clinics and community mental health centers that are chronically underresourced, understaffed, and overextended [4,5]. These public sector agencies are rarely able to meet the demand for services [6]. Barriers in the capacity to provide high-quality care on the provider's side may be compounded by hesitancy around treatment seeking on the patient's side [7]. People with SMI are often exposed to pervasive societal stigma about their conditions and therefore can be reluctant to openly seek services at a clinic if it risks them being labeled *mentally ill* [8,9]. The consequence of these intersecting challenges is a vicious cycle in which those who are most impaired receive the least amount of support, thus deteriorating even more over time.

People with SMI are not dramatically different from the general population in their use of mobile technologies [10,11]. Owing to subsidy programs that offer people with disabilities access to mobile phones and data plans (eg, Federal Communications Commission Lifeline Program) and the dropping prices of mobile phones worldwide, people with SMI now have unprecedented opportunities to access information, stay connected to others, and potentially receive services remotely via personal mobile devices [12]. Survey studies conducted in recent years have shown that approximately two-thirds of people with SMI in the United States already own smartphones and that most are interested in leveraging them to support their health care [13-16]. Academic research groups and commercial companies are responding to the opportunity and have begun to develop and test a range of smartphone-supported digital health technologies for people with SMI, including illness self-management apps, virtual peer support platforms, and sensor-enabled remote patient monitoring systems [17-19].

Conducting the research necessary to demonstrate the safety and clinical utility of novel smartphone interventions for people with SMI is challenging and costly. Recruitment of participants for clinical trials is the most salient cause of study delays and a major obstacle in expeditiously moving novel and potentially helpful interventions to real-world practice [20]. Traditional participant recruitment strategies are dependent on collaborative partnerships with study sites and clinician referrals [21]. This creates potential biases in clinical trial samples, as the enrolled individuals already need to be known by health care providers and be actively receiving care to be offered the opportunity to engage in research. Therefore, the participant samples used in standard clinical trials may be particularly unrepresentative of hard-to-reach populations, such as people with SMI, who are known to disengage from clinic-based services for extended periods [22,23].

To overcome these obstacles, a growing number of studies have used web-based participant recruitment strategies, including social media advertisements, virtual outreach through web-based interest groups, search engine advertisements, and various other website campaigns. Across clinical populations, these efforts have yielded impressive results in terms of cost-effectiveness, time efficiency, and reach [24]. In clinical trials where the experimental intervention is digital and behavioral in nature (ie, does not require direct contact between a patient and a provider) and the comparator condition does not require in-person contact with the research team (eg, no intervention control, waitlist control, or digital attention–control alternative), recruitment, treatment, and assessment of outcomes can all be conducted fully remotely and potentially on a single device [25-30]. To our knowledge, fully remote randomized controlled trials have not been conducted with people with SMI [24].

Objective

Here, we report the first fully remote randomized controlled clinical trial of a mobile health (mHealth) intervention for people with SMI. The first objective of this study is to evaluate whether individuals with SMI can be successfully recruited, assessed, and engaged in a digital intervention in a fully remote clinical trial. The second objective of the study is to evaluate the clinical effectiveness of the CORE intervention, a smartphone app designed to challenge dysfunctional thoughts that underlie common symptoms of SMI, self-stigmatizing attitudes, and maladaptive beliefs that impede treatment seeking and recovery.

Methods

Study Design

This study involved a fully remote randomized controlled crossover waitlist trial design. All participants were given the opportunity to receive the study intervention. Following enrollment, participants were randomized to either receiving the CORE intervention (indicated as *1*) or being in the waitlist control group (indicated as *0*) in blocks of 4. The group allocation for each block was selected randomly from a list of 6 possible sequences (ie, 1100, 1010, 1001, 0110, 0101, and

0011). In the active intervention group, participants completed the baseline assessment and were then immediately given access to the CORE app for 30 days of use. After a month, they concluded the intervention, uninstalled the app, and completed a second assessment. After an additional month (at 60 days), they completed a third assessment to measure the stability of symptom change post intervention. In the control arm, participants waited 30 days to receive the CORE app. After a month, they completed a second assessment and were provided access to the CORE app. After an additional month (at 60 days), they completed a third assessment to measure within-subject changes.

Procedures

All study procedures were reviewed and approved by the institutional review board of the University of Washington (ID number 00006898). We put a detailed and institutional review board–approved plan in place to respond to cases of increased risk. Web-based recruitment was conducted through advertisements on Google and Facebook. Google advertisements are presented depending on the user's search terms. For this study, we selected a range of terms associated with severe mental illness (eg, *schizophrenia*, *bipolar*, *seeing things*, and *am I crazy*) and related keywords generated by the Google *broad match* algorithm. Facebook advertisements use a similar methodology to target individual interests. For this study, we targeted interests such as mental health and schizophrenia. Individuals who clicked on the advertisements were directed to the study website. The study website provided written and video descriptions of the project, a downloadable version of the study consent form, and an option to complete a screening questionnaire via a *see if I am eligible* button. If eligible, participants were again presented with the consent form and were required to answer questions demonstrating their understanding of the study details. If they answered these questions correctly within 3 attempts, they could proceed to the baseline assessments. Participants were excluded if they were unable to complete this step successfully, as they were deemed unable to provide informed consent. Participants completed a battery of self-report questionnaires on the web at baseline. If participants' responses to questionnaires indicated the presence of suicidal ideation, they were immediately provided with resources for emergency support. Upon completion of baseline questionnaires, participants were randomized to either an active intervention or waitlist control group. Those randomized to the active intervention arm were given instructions on how to download and install the CORE mobile app.

At the start of the study period, staff contacted participants via email or SMS text messaging to welcome them to the study and remind them to complete the follow-up assessments at 30 and 60 days. After this welcome message, participants interacted with the study team only if they reached out to the study staff for technical support. When participants were due for an assessment, they were sent an automated SMS text message to alert them to complete the battery through a weblink. Participants were incentivized by being entered into lotteries to win US $500 and US $1000 after completing the 30- and 60-day assessments, respectively.

A member (JT) of our research team who was blinded to participant allocation completed all study analyses, and these analyses were determined beforehand. All outcome measures were preselected by the study team and examined in this study. The sample size was calculated to fit analytic requirements, and study recruitment was discontinued once the required sample size was met.

Participants

Participants were eligible if they could speak English; were aged >18 years; lived in the United States; self-reported a diagnosis of schizophrenia or schizoaffective disorder, bipolar disorder, or major depressive disorder; and owned a smartphone with a data plan. They were excluded if they had previously participated in the study or were unavailable for 60 days of participation.

Assessments

At all assessments, we administered measures of depression (Beck Depression Inventory–II [BDI-II] [31]), anxiety (Generalized Anxiety Disorder–7 [GAD-7] [32]), auditory hallucinations (Hamilton Program for Schizophrenia Voices Questionnaire [HPSVQ] [33]), paranoid thinking (Green Paranoid Thoughts Scale [GPTS] [34]), recovery (Recovery Assessment Scale [RAS] [35,36]), self-esteem (Rosenberg Self-Esteem Scale [RSES] [37]), social isolation (The Friendship Scale [38]), and functional disability (Sheehan Disability Scale [SDS] [39]). At baseline, we collected information from participants regarding demographics and technology use. After the intervention period, participants also completed a 26-item self-report usability and acceptability measure comprising adapted items from the System Usability Scale [40], Post Study System Usability Questionnaire [40], Technology Assessment Model Measurement Scales [41], and Usefulness, Satisfaction, and Ease questionnaire [42]. Participants were asked to rate their agreement with a series of statements about the intervention. A similar measure was used in previous studies conducted by our group [43].

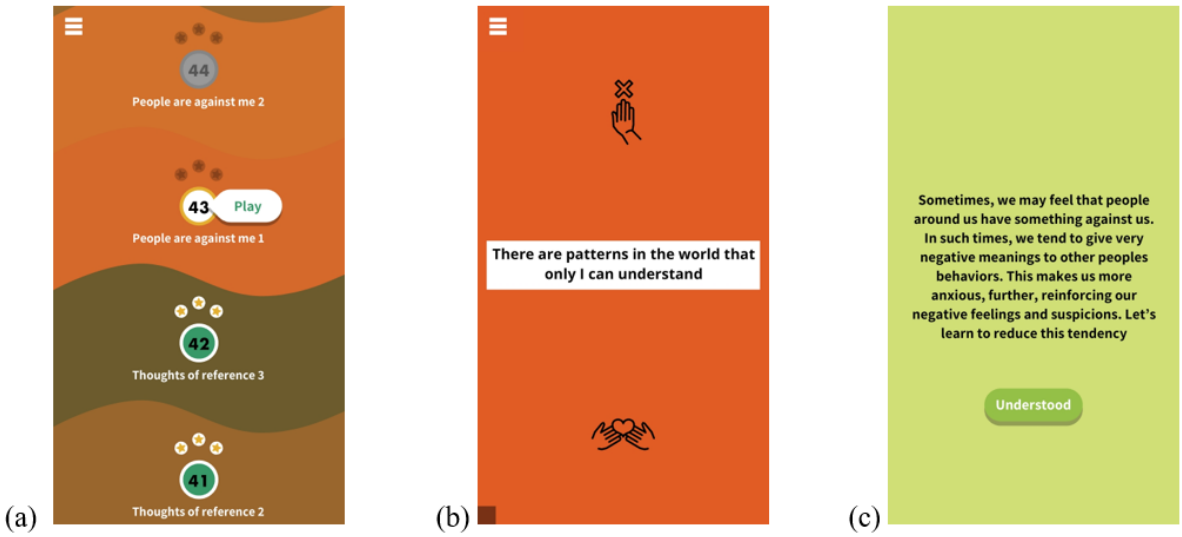
Intervention Description

The CORE intervention is a smartphone app that uses the GGtude platform, a system designed to increase the cognitive flexibility of individuals struggling with a range of mental health problems through brief daily training [44-47]. CORE was specifically designed to help counteract dysfunctional thoughts in multiple domains that are relevant to the subjective experience of having SMI. The intervention comprises daily brief game-like exercises designed to produce changes in the relative activation of adaptive and maladaptive beliefs about the self, others, and the world such that adaptive beliefs would be more easily retrieved than maladaptive ones.

CORE users were trained to respond to multiple statements in a sequence of modules that progress through the following domains: *self-talk*, *belief in change*, *self-stigma*, *self-care*, *self-worth*, *illness and identity*, *personal strength*, *social avoidance*, *feelings versus facts*, *catastrophization*, *thoughts of reference*, *paranoid ideation* (ie, *people are against me*), *treatment seeking*, and *recovery*. The modules begin with brief psychoeducation about the target domain and how maladaptive

beliefs can hamper recovery. Thoughts appear as statements on the smartphone screen, and users are required to either endorse (ie, *drag* the statement down toward them on the touchscreen) or discard them (ie, *push* the statement upwards away from them). Users learn to embrace self-statements reflecting more nuanced adaptive thoughts (eg, belief in change, importance of self-care, alternative explanations to threat perceptions, and value of treatment seeking). The module content was organized into 53 levels, and participants were recommended to not complete more than 3 levels a day for a period of 30 days. Push notifications reminded users to complete their daily training. After each level, participants had the option of adding to their personal toolbox 1 of 3 positive statements that they most related to. Participants could access their toolbox at any time. Figure 1 depicts screenshots of the targeted beliefs menu, an example of a maladaptive statement, and a psychoeducational element.

Figure 1. Screenshots of the (a) targeted beliefs menu, (b) an example of a maladaptive statement, and (c) a psychoeducational element.



Data Analysis

Means, SDs, and frequencies were reported using descriptive statistics for all participant demographics, symptom measures, and acceptability and usability scale items. A series of *t* tests and chi-square tests were used to test for group differences in demographic variables and symptom measures at baseline (T1) before examining intervention effects at 30 days (T2) and 60 days (T3). Pre- and postscores for both groups were explored using a series of repeated-measures analysis of variance (ANOVA). The intervention effect between groups at T2 was initially tested using a 2×2 mixed-design repeated-measures ANOVA, controlling for participant diagnosis and baseline assessment scores. A 1×3 repeated measures ANOVA tested the performance of the active group across T1, T2, and T3, with post hoc Bonferroni pairwise comparisons used to explore outcomes when significant differences occurred. Finally, a series of 1×3 repeated measures ANOVAs tested the intervention effect after crossover (T2 and T3) for the waitlist control group participants to see if intervention effects were replicated. A series of equivalence tests were conducted using independent sample 2-tailed *t* tests to determine whether the magnitude of intervention effects differed for the active group participants between T1 and T2 as compared with the intervention effect for waitlist control participants between T2 and T3. Spaghetti plots were provided to visualize changes over time across symptom measures in both groups. To minimize bias during analysis, we used an intent-to-treat approach and implemented a multiple imputation strategy to replace missing values [48]. Multiple imputation is considered the most appropriate method of handling missing data for a study of this size, has been used successfully in repeated measure designs, and has been shown to have utility with levels of missing data greater than that observed in this study [49,50]. Sensitivity analyses using pattern-mixture models were used to establish data that were missing at random before imputation [51]. Our imputation model was specified using data from participants in each condition and included participant demographic variables (ie, diagnosis, gender, age, and race) and all baseline assessment scores.

Results

Recruitment and Enrollment

Web-based recruitment advertisements were placed between January 2020 and September 2020, using a total budget of US $2984.26. We recruited a total of 1123 people from these advertisements; of these 1123 people, 315 (28.04%) were randomized for the trial, and 808 (71.95%) were excluded. Of the 315 participants, 154 (48.9%) dropped out of the study before T2 and 45 (14.3%) dropped out between T2 and T3. Approximately 41.5% (64/154) of participants in the active group and 60.2% (97/161) of participants in the waitlist group were retained at T2, and 33.1% (51/154) of participants in the active group and 40.3% (65/161) of individuals in the waitlist group were retained at T3. Those who dropped out of the study were not found to differ significantly from those who were retained in terms of demographic characteristics or baseline scores on outcome measures (Figure 2). Most participants were recruited from Google advertisements (226/315, 71.7%). Overall, Google advertisements were viewed 557,700 times

and were ultimately clicked on 5100 times. The states with the most clicks on the Google advertisement were California (540/5100, 10.59%), Texas (393/5100, 7.71%), Florida (295/5100, 5.78%), and New York (294/5100, 5.77%). The most successful Google advertisement keywords for increasing engagement from potential participants were *mental health*, *mental illness*, *depression*, *illuminati*, and *bipolar*. Most individuals (4734/5100, 92.82%) clicked on the advertisement through mobile phones, whereas few used desktop computers (256/5100, 5.02%) or tablets (110/5100, 2.16%). Participants that enrolled in the study were recruited from 45 states (Figure 3), with the most participants coming from California (24/315, 7.6%), Texas (21/315, 6.7%), Florida (17/315, 5.4%), New York (16/315, 5.1%), Pennsylvania (16/315, 5.1%), and Washington (16/315, 5.1%). We successfully enrolled participants from 80% (12/15) of the most rural states in the country.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram.

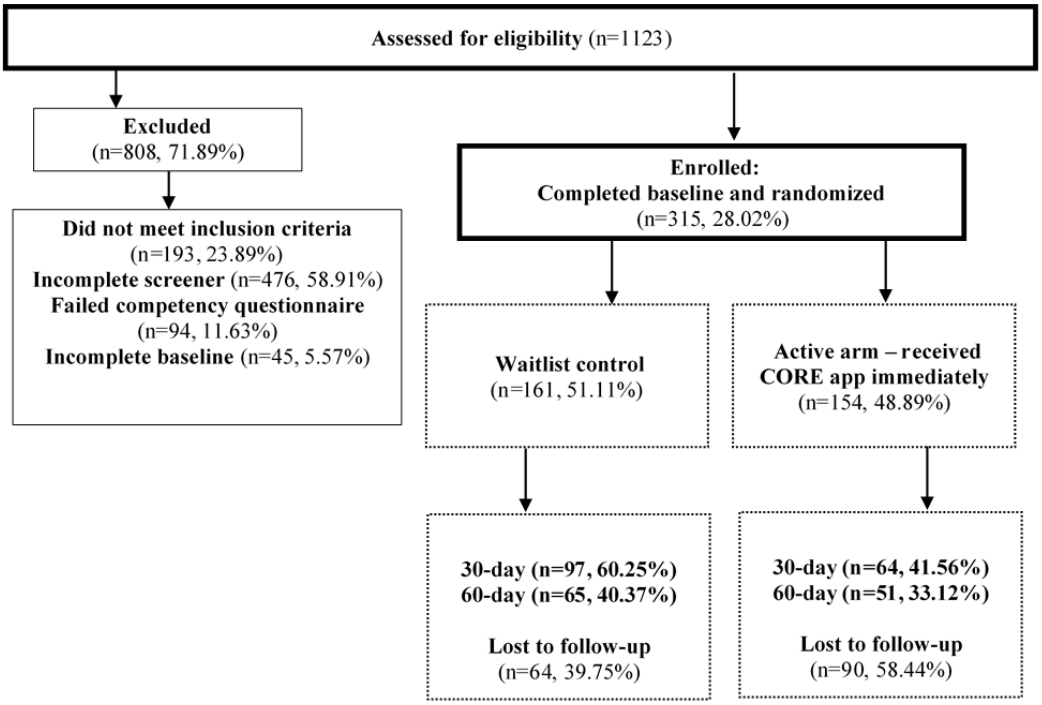
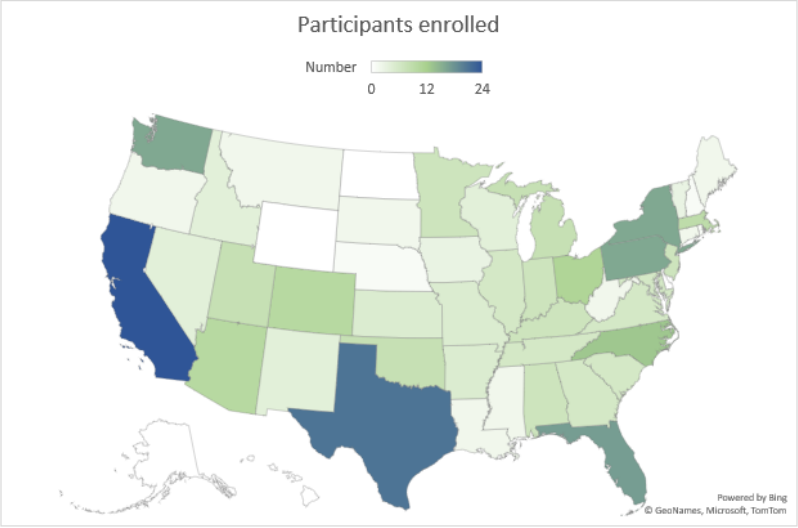


Figure 3. Map of participants enrolled based on location.



Demographics

Full demographics of the study participants are reported in Table 1. The final study participants were mostly female (264/315, 83.8%), White (241/315, 76.5%), heterosexual (218/315, 69.2%), living independently (150/315, 47.6%) or with family (136/315, 43.2%), unemployed (210/315, 66.7%), and aged between 18-78 years. No demographic differences were observed following randomization between the 2 study groups.

Table 1. Participant demographics and clinical history at baseline (N=315).

Characteristics	Waitlist (n=161)	Active (n=154)
Age (years), mean (SD)	36.85 (10.91)	38.98 (12.30)
Gender, n (%)		
Female	132 (82)	132 (85.7)
Male	19 (11.8)	18 (11.7)
Transgender	6 (3.7)	0
MTF[a]	2 (1.2)	0
FTM[b]	4 (2.5)	0 (0)
Nonbinary	4 (2.5)	2 (1.3)
Other	0	2 (1.3)
Diagnosis, n (%)		
Bipolar disorder	61 (37.9)	50 (32.5)
Major depressive disorder	67 (41.6)	69 (44.8)
Schizophrenia or schizoaffective disorder	33 (20.5)	35 (22.7)
Race, n (%)		
White	118 (73.3)	123 (79.9)
Black or African American	20 (12.4)	12 (7.8)
Asian	6 (3.7)	4 (2.6)
Alaskan Native or American Indian	1 (0.6)	1 (0.6)
Declined to answer	0	1 (0.6)
More than one race	16 (9.9)	13 (8.4)
Ethnicity, n (%)		
Spanish, Hispanic, or Latino	17 (10.6)	13 (8.4)
Not Spanish, Hispanic, or Latino	143 (89.4)	139 (91.4)
Education, n (%)		
Finished eighth grade	3 (1.9)	1 (0.7)
Some high school	10 (6.2)	5 (3.3)
High school diploma	31 (19.3)	22 (14.4)
Some college or technical school	47 (41.6)	62 (40.5)
Associate’s degree	15 (9.3)	21 (13.7)
Bachelor’s degree	26 (16.1)	22 (14.4)
Some graduate school	1 (0.6)	6 (3.9)
Master’s degree	7 (4.3)	13 (8.5)
Doctorate	1 (0.6)	1 (0.7)
Employment status, n (%)		
Full-time	23 (14.3)	29 (19)
Part-time	27 (16.8)	17 (11.1)
Unemployed	105 (65.2)	105 (68.6)
Living situation, n (%)		
Independent	78 (48.4)	72 (47.1)
Living with family	71 (44.1)	65 (42.5)
Homeless	6 (3.7)	10 (6.5)

Characteristics	Waitlist (n=161)	Active (n=154)
Substance use treatment	0	1 (0.7)
Assisted or supported living	6 (3.7)	5 (3.3)
Lifetime psychiatric hospitalizations, n (%)		
0	49 (30.4)	44 (28.8)
1-5	69 (42.9)	72 (47.1)
6-10	20 (12.4)	18 (11.8)
11-15	11 (6.8)	5 (3.3)
16-20	3 (1.9)	3 (2)
>20	9 (5.6)	11 (7.2)
Past-year psychiatric hospitalizations, n (%)		
0	127 (78.9)	114 (74.5)
1-5	33 (20.5)	37 (24.2)
6-10	1 (0.6)	2 (1.3)
Frequency of auditory hallucinations, n (%)		
Never	82 (51.9)	75 (49)
A few times a year	24 (15.2)	23 (15)
Once or twice a month	17 (10.8)	15 (9.8)
Once a week	5 (3.2)	9 (5.9)
≥2 times a week	30 (19)	31 (20.3)
Beck Depression Inventory-II, mean (SD)	34.57 (13.34)	33.46 (13.75)
Green Paranoid Thought Scale, mean (SD)	88.24 (35.63)	86.79 (36.01)
Hamilton Program for Schizophrenia Voices, mean (SD)	19.86 (9.03)	18.42 (8.81)
Generalized Anxiety Disorder–7, mean (SD)	13.89 (5.77)	14.42 (5.16)
Sheehan Disability Scale, mean (SD)	24.11 (7.21)	24.11 (7.02)

[a]MTF: male-to-female.

[b]FTM: female-to-male.

Clinical Characteristics

The sample comprised individuals with bipolar disorder (111/315, 35.2%), major depressive disorder (136/315, 43.2%), and schizophrenia or schizoaffective disorder (68/315, 21.6%). Approximately a quarter (75/315, 23.8%) of the participants reported experiencing auditory hallucinations on a weekly or more frequent basis at baseline. Approximately 23.2% (73/315) of participants reported experiencing a psychiatric hospitalization in the past year, and 25.4% (80/315) of participants reported having ≥6 psychiatric hospitalizations in their lifetime. Baseline scores on clinical outcome measures were in the severe range on both the BDI-II (mean 33.95, SD 13.61) and the GPTS (mean 87.51, SD 35.79) and in the moderate range on the GAD-7 (mean 14.12, SD 5.48) and HPSVQ (mean 19.08, SD 8.89). Participants had a moderate level of disability SDS (mean 24.07, SD 7.12). Significant correlations existed between scores on the BDI-II and GAD-7 (r=0.69; P<.001), BDI-II and GPTS (r=0.38; P<.001), and GAD-7 and GPTS (r=0.40; P<.001), indicating a high degree of comorbidity within the sample. No differences were found between groups on any clinical characteristics or baseline outcome measures after randomization. Diagnosis was not found to be a significant predictor of completion at T2 or T3, with the distribution of completion resembling the distribution of participants at baseline. At T2, 20.5% (33/161) of completers had a diagnosis of schizophrenia or schizoaffective disorder, 44.1% (71/161) had a diagnosis of major depressive disorder, and 35.4% (57/161) had a diagnosis of bipolar disorder. At T3, 23.3% (27/116) of completers had a diagnosis of schizophrenia or schizoaffective disorder, 46.6% (54/116) had a diagnosis of major depressive disorder, and 30.2% (35/116) had a diagnosis of bipolar disorder.

Engagement Metrics

We were able to collect the CORE app use data from 51.4% (162/315) of participants. Of the 162 participants, 82 (50.6%) completed all 53 intervention levels, whereas participants, on average, completed 35 levels. Most participants reported that they would like to use CORE more often; that if they had access to CORE, they would use it; that the app was easy to use and sufficiently interactive; and that they did not need technical support to use CORE. Participants rated that they were satisfied with the CORE intervention and that they would recommend

it to a friend. The distribution of participants' responses to all usability and acceptability questions is displayed in Table 2.

Table 2. Participant usability and acceptability ratings (N=119).

Item	Disagree, n (%)	Neutral, n (%)	Agree, n (%)
I think that I would like to use CORE often.	8 (6.7)	26 (21.8)	85 (71.4)
I thought CORE was easy to use.	1 (0.8)	5 (4.2)	113 (95.0)
I found that the different parts of CORE work well together.	1 (0.8)	23 (19.3)	95 (79.8)
I would imagine that most people would learn to use CORE very quickly.	0 (0.0)	6 (5.0)	113 (95.0)
I felt very confident using CORE.	1 (0.8)	24 (20.2)	94 (79.0)
Overall, I am satisfied with how easy it is to use CORE.	1 (0.8)	8 (6.7)	110 (92.4)
I was able to use the *modules* quickly in CORE.	3 (2.5)	8 (6.7)	108 (90.8)
I felt comfortable using CORE.	0 (0.0)	9 (7.6)	110 (92.4)
It was easy to learn to use CORE.	0 (0.0)	7 (5.9)	111 (94.1)
Whenever I made a mistake using CORE, I could recover easily and quickly.	2 (1.7)	17 (14.3)	100 (84.0)
It was easy to find the information I needed.	5 (4.2)	16 (13.6)	97 (82.2)
The information provided for CORE was easy to understand.	2 (1.7)	10 (8.5)	105 (89.7)
How things appeared on the screen was clear.	1 (0.8)	9 (7.6)	108 (91.5)
If I have access to CORE, I will use it.	6 (5.0)	17 (14.3)	96 (80.7)
I am satisfied with CORE.	7 (5.9)	11 (9.2)	101 (84.9)
I would recommend CORE to a friend.	6 (5.0)	23 (19.3)	90 (75.6)
CORE is fun to use.	7 (5.9)	44 (37.0)	68 (57.1)
CORE works the way I want it to work.	10 (8.4)	28 (23.5)	81 (68.1)
I feel I need to have CORE.	20 (16.8)	58 (48.7)	41 (34.5)
CORE helped me manage my symptoms.	12 (10.1)	41 (34.5)	66 (55.5)
CORE was interactive enough.	6 (5.0)	26 (21.8)	87 (73.1)
I found CORE to be very complicated	108 (90.8)	7 (5.9)	4 (3.4)
I think that I would need the support of a technical person to be able to use CORE.	103 (86.6)	7 (5.9)	9 (7.6)
I thought there was too much inconsistency in CORE.	99 (83.2)	13 (10.9)	7 (5.9)
I found CORE very awkward to use.	103 (86.6)	11 (9.2)	5 (4.2)
I needed to learn a lot of things before I could get going with CORE.	105 (88.2)	8 (6.7)	6 (5.0)

Between-Group Differences: Active Group Versus Waitlist Control Group

Analyses of differences between the active group and the waitlist control group from T1 to T2 revealed a significant treatment×time interaction effect for the BDI-II ($F_{1,313}$=13.38; P<.001), GAD-7 ($F_{1,313}$=5.87; P=.02), RAS ($F_{1,313}$=23.42; P<.001), RSES ($F_{1,313}$=19.28; P<.001), and SDS ($F_{1,313}$=10.73; P=.001). Large effects were observed at T2 for the BDI-II (d=0.58), RAS (d=0.61), and RSES (d=0.64). A moderate effect was observed for the SDS (d=0.44) and a small effect for the GAD-7 (d=0.20).

This indicates that participants engaging in the active condition showed improvements on these outcome measures after 30 days of using the app compared with participants in the waitlist control group who did not have access to the app over the same period (Figure 4). A significant main effect of time was found for the BDI-II ($F_{1,313}$=44.33; P<.001), GAD-7 ($F_{1,313}$=66.77; P<.001), GPTS ($F_{1,313}$=39.85; P<.001), RAS ($F_{1,313}$=46.15; P<.001), RSES ($F_{1,313}$=28.89; P<.001), Friendship Scale ($F_{1,313}$=21.05; P<.001), and SDS ($F_{1,313}$=51.644; P<.001). Table 3 displays the means and SDs for all measures at each time point for the active group participants and the waitlist control group participants.

Figure 4. Outcome measures (y-axis) at baseline (T1), 30 days (T2), and 60 days (T3) assessments points (x-axis) for active group and waitlist control group. BDI-II: Beck Depression Inventory-Second Edition; GAD-7: Generalized Anxiety Disorder-7; GPTS: Green Paranoid Thought Scale; HPSVQ: Hamilton Program for Schizophrenia Voices; RAS: Recovery Assessment Scale; RSES: Rosenberg Self-esteem Scale; SDS: Sheehan Disability Scale.

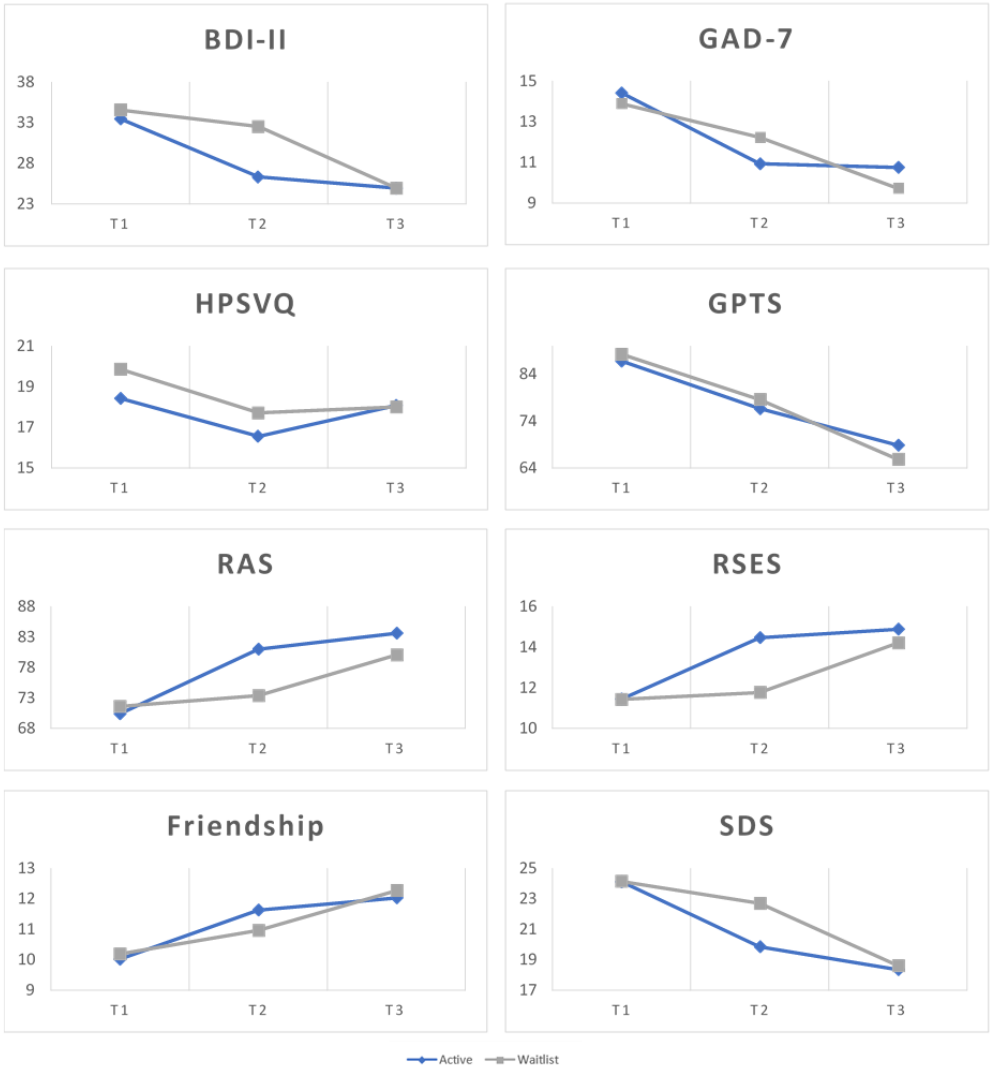


Table 3. Comparison of outcome measures across periods for each group.

Measure	Baseline, mean (SD)	30 days, mean (SD)	60 days, mean (SD)
Beck Depression Inventory–II			
Active	33.46 (13.75)	26.33 (10.43)[a,b]	24.94 (8.71)
Waitlist	34.57 (13.34)	32.52 (11.12)	24.93 (10.59)[b]
Generalized Anxiety Disorder-7			
Active	14.42 (5.16)	10.93 (4.17)[b,c]	10.75 (3.69)
Waitlist	13.89 (5.77)	12.23 (5.00)[d]	9.72 (4.53)[b]
Hamilton Program for Schizophrenia Voices Questionnaire			
Active	18.42 (8.81)	16.56 (8.54)	18.08 (6.07)
Waitlist	19.86 (9.03)	17.71 (10.20)	18.00 (11.99)
Green Paranoid Thoughts Scale			
Active	86.79 (36.01)	76.66 (23.97)[b]	68.84 (20.41)[b]
Waitlist	88.24 (35.63)	78.59 (27.54)[b]	65.84 (24.11)[b]
Recovery Assessment Scale			
Active	70.42 (16.62)	81.02 (11.01)[a,b]	83.64 (9.44)[b]
Waitlist	71.61 (17.18)	73.38 (12.44)	80.06 (13.08)[b]
Rosenberg Self-Esteem Scale			
Active	11.45 (6.39)	14.47 (4.28)[a,d]	14.89 (4.28)
Waitlist	11.43 (5.83)	11.77 (4.96)	14.21 (5.01)[b]
Friendship scale			
Active	10.02 (3.84)	11.63 (3.19)[b]	12.03 (3.00)
Waitlist	10.19 (3.56)	10.96 (3.39)	12.27 (3.28)[b]
Sheehan Disability Scale			
Active	24.12 (7.03)	19.85 (6.26)[b,e]	18.35 (6.38)[d]
Waitlist	24.14 (7.21)	22.70 (5.83)[d]	18.61 (6.54)[b]

[a]Significance of between-group change within period (active vs waitlist control) <.001.

[b]Significance of within-group change from the previous period <.001.

[c]Significance of between-group change within period (active vs waitlist control) <.05.

[d]Significance of within-group change from the previous period <.01.

[e]Significance of between-group change within period (active vs waitlist control) <.01.

Within-Group Change: Active Intervention Group at T2 and Maintenance at T3

A series of 1×3 repeated-measures ANOVA was used to examine intervention effects within the active group between T1 and T2 and whether these effects were maintained 30 days after discontinuation of app use (Table 3). Results of the ANOVAs showed significant within-group differences for the BDI-II ($F_{2,306}$=43.59; P<.001), GAD-7 ($F_{2,306}$=43.71; P<.001), GPTS ($F_{2,306}$=34.39; P<.001), RAS ($F_{2,306}$=65.22; P<.001), RSES ($F_{2,306}$=32.85; P<.001), Friendship Scale ($F_{2,306}$=17.94; P<.001), and SDS ($F_{2,306}$=56.21; P<.001). Large effect sizes were observed between T1 and T2 for the active intervention group on the BDI-II (d=0.60), GAD-7 (d=0.69), RAS (d=0.70), and SDS (d=0.63). Moderate to small effect sizes occurred during this period on the RSES (d=0.47), GPTS (d=0.35), and Friendship Scale (d=0.39). No significant change was found at any time point for the active group for the HPSVQ ($F_{2,306}$=0.704; P=.51). Bonferroni post hoc comparisons revealed significant improvements in each outcome measure at T2 except for the HPSVQ. All improvements were maintained between T2 and T3, with an additional small effect noted on the GPTS (d=0.33), RAS (d=0.27), and SDS (d=0.24), suggesting a prolonged intervention effect.

Within-Group Change: Waitlist Control Group at T2 and Following Crossover at T-3

Results from the active group were replicated following crossover in the waitlist control group between T2 and T3. Participants in the waitlist control group were found to have significant differences on the BDI-II ($F_{2, 322}$=60.79; P<.001), GAD-7 ($F_{2, 322}$=50.41; P<.001), GPTS ($F_{2, 322}$=55.15; P<.001), RAS ($F_{2, 322}$=32.03; P<.001), RSES ($F_{2, 322}$=34.35; P<.001), Friendship Scale ($F_{2, 322}$=26.28; P<.001), and SDS ($F_{2, 322}$=50.93; P<.001). The lack of change in the HPSVQ scores after 30 days of CORE use was also replicated ($F_{2, 322}$=1.14; P=.34). Post hoc analysis revealed that significant improvements took place for the waitlist control group between T2 and T3 for all outcome measures except HPSVQ. Effect size magnitudes were similar to those seen in the active group during the previous period, with large to moderate effects on the BDI-II (d=0.73), GAD-7 (d=0.46), GPTS (d=0.48), RAS (d=0.55), RSES (d=0.52), Friendship Scale (d=0.43), and SDS (d=0.67). Improvements in the waitlist control group were also found between T1 and T2 for GAD-7, GPTS, and SDS. These improvements as a result of time were sufficient to negate a between-subject effect for the GPTS but not for the BDI-II or the GAD-7. The magnitude of change observed after 30 days of CORE use between groups was equivalent across all outcome metrics except for the RAS (t_{105}=2.28; P=.03), with the active group showing a slightly larger improvement than the waitlist control group.

Discussion

Principal Findings

Recent studies have demonstrated the feasibility of conducting remote trials of digital health apps for common mental health problems, including stress, anxiety, and depression [27-29]. This paper reports on the first fully remote randomized controlled trial of a smartphone intervention involving people with more severe forms of psychopathology, including schizophrenia and bipolar disorder. Building on and expanding web-based enrollment strategies used in previous works [52,53], we were able to reach, recruit, randomize, treat, and assess a sociodemographically diverse sample of participants from 45 states, far exceeding the reach of traditional localized study recruitment approaches and at a fraction of the cost. A key advantage of remote research strategies leveraging web-based recruitment advertisements lies in the flexibility they afford the investigative team to determine the geographical regions and roll out of study recruitment materials. Similar to what has been shown with other clinical populations [24], this proved to be a remarkably efficient methodology of reaching people with SMI who are more impaired and often considered more complex to engage in research [54].

This study found that a novel smartphone app, CORE, proved to be usable, acceptable, and effective in improving recovery and reducing the severity of psychiatric symptoms among individuals with SMI. Our findings align with an ample body of digital mental health research suggesting that mHealth smartphone apps can be of clinical value to people with SMI [17,55,56]. Integration of technology with human support can bolster the engagement and clinical potency of technology-based treatments for SMI [57,58]. The CORE intervention was entirely self-navigated by users in the study who received minimal or no remote technical support from our study staff. We are not advocating for a shift toward fully automated approaches as the preferred model. However, what this study did demonstrate is that when technology is designed with the characteristics of the intended users in mind (ie, in terms of functionality, accessibility, navigability, and content), deployment of specialty digital mental health tools that do not involve humans in the loop of care can also produce significant clinical benefits.

The CORE intervention builds on a digital cognitive training strategy that has been previously demonstrated to reduce dysfunctional *self-talk* and increase resilience in community-based subclinical groups [44-46]. The app included modules focused on countering maladaptive beliefs in common mental health domains (eg, self-esteem, distinguishing thoughts from feelings, social anxiety, and catastrophizing) as well as modules specifically designed for people with SMI (eg, stigma related to mental illness, threat perception and persecutory ideation, hopelessness, strengths-based recovery, self-care, and treatment seeking). Thus, it is meant to be used transdiagnostically [59].

Although this study was a fully remote trial that relied entirely on participants' self-reports, we are confident that we were able to reach our intended target audience; most of our study sample reported having a psychiatric diagnosis of schizophrenia, schizoaffective disorder, or bipolar disorder; a history of multiple psychiatric hospitalizations; and being unemployed. Approximately one-tenth reported being currently homeless or residing in an assisted-living facility. Combined with the sample's average baseline ratings showing moderate to severe psychiatric symptom severity and moderate to high levels of disability, our data suggest that the sample comprised people with significant functional impairment.

The waitlist-controlled study design enabled the evaluation of both between-group effects during the intervention period and within-group changes in both groups over time. The results demonstrated a link between the timing of participants' exposure to the treatment app and when they experienced significant changes in key clinical and functional measures. A combination of factors such as simplicity, brevity, daily use, and game-like interactions may have encouraged the use of the CORE app. Daily practice comprising identification and categorization of self-statements, repeated exposure to adaptive self-statements, and psychoeducation may have facilitated retrieval of adaptive beliefs over maladaptive ones, thereby reducing the severity of symptoms [60]

This study has several limitations. The opportunity to participate was available only to individuals who used the internet (ie, to receive study recruitment advertisements) and who owned smartphones (ie, to download the CORE app) and therefore may not be representative of the full range of people with SMI. Similar to fully remote digital mental health trials involving participants with less severe psychopathology [61], our study sample was overrepresented by female participants (264/315,

83.8%), and study dropout rates were high, requiring significant data imputation. All screening and assessment questions relied on individuals' self-reports but were not corroborated by a trained clinician or additional documentation (eg, EHR data). Replication of our findings in samples that were screened, diagnosed, and assessed by trained clinical assessors could bolster confidence in the nature of our results. Using a waitlist control is often considered a more ethical study design than offering no treatment at all or sham interventions. However, it may also have artificially inflated estimates of the intervention effects; asking participants who are ready for treatment to wait to receive CORE may have frustrated them, stalled their attempts to make independent changes, or artificially delayed their seeking other treatment options. Finally, although our measure of recovery taps *willingness to ask for help* [35,36], we did not formally evaluate changes in treatment-seeking behavior, which is an area of focus in the CORE training program and the key variable of interest.

Conclusions

Mental health researchers, funders, industry leaders, patients, and their caregivers have been advocating for the development and deployment of effective digital health tools to improve the outcomes of people with psychiatric conditions [62,63]. The global COVID-19 pandemic has led to major disruptions in the delivery of standard mental health services and has shed light on the vulnerabilities intrinsic to complete reliance on clinic-based treatment models [64,65]. In the context of this ongoing public health crisis, new scientific evidence showing that remotely-accessed mHealth technologies such as CORE can be navigable and beneficial to people with SMI is very encouraging. Regulatory bodies have taken active steps to remove barriers to the use of digital health technologies for psychiatric disorders [66]. Currently, rapid adoption and real-world dissemination of evidence-based digital health interventions are needed if we are to shorten the science-to-service gap and help address the significant unmet mental health needs of people with SMI during the pandemic and beyond.

Acknowledgments

The authors DBZ, AC, JT, SN, and AC were supported by an award from National Institute of Mental Health (NIMH) (R01MH112641). BB was supported by a Mentored Patient-Oriented Research Career Development Award from NIMH (K23MH122504).

Conflicts of Interest

DBZ has an intervention content licensing agreement with Pear Therapeutics and has a financial interest in Merlin LLC and FOCUS technology. He has consulted with Trusst Health, K Health, eQuility, and Otsuka Pharmaceuticals Ltd. GD is the cofounder of GGtude Ltd and has a financial interest in the CORE app described in this paper. Data analyses were conducted by members of the team who were unaffiliated with GGtude Ltd. All the remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 370 KB - jmir_v23i11e29201_app1.pdf]

References

1. Harvey PD, Heaton RK, Carpenter WT, Green MF, Gold JM, Schoenbaum M. Functional impairment in people with schizophrenia: focus on employability and eligibility for disability compensation. Schizophr Res 2012 Sep;140(1-3):1-8 [FREE Full text] [doi: 10.1016/j.schres.2012.03.025] [Medline: 22503642]
2. Folsom DP, Hawthorne W, Lindamer L, Gilmer T, Bailey A, Golshan S, et al. Prevalence and risk factors for homelessness and utilization of mental health services among 10,340 patients with serious mental illness in a large public mental health system. Am J Psychiatry 2005 Feb;162(2):370-376. [doi: 10.1176/appi.ajp.162.2.370] [Medline: 15677603]
3. Sylvestre J, Notten G, Kerman N, Polillo A, Czechowki K. Poverty and serious mental illness: toward action on a seemingly intractable problem. Am J Community Psychol 2018 Mar;61(1-2):153-165. [doi: 10.1002/ajcp.12211] [Medline: 29243829]
4. Rosenblatt RA, Andrilla CH, Curtin T, Hart LG. Shortages of medical personnel at community health centers: implications for planned expansion. JAMA 2006 Mar 01;295(9):1042-1049. [doi: 10.1001/jama.295.9.1042] [Medline: 16507805]
5. Rosenberg S, Rosenberg J. Community Mental Health: Challenges for the 21st Century. New York: Routledge; 2017.
6. Mojtabai R, Fochtmann L, Chang S, Kotov R, Craig TJ, Bromet E. Unmet need for mental health care in schizophrenia: an overview of literature and new data from a first-admission study. Schizophr Bull 2009 Jul;35(4):679-695 [FREE Full text] [doi: 10.1093/schbul/sbp045] [Medline: 19505994]
7. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
8. Ben-Zeev D, Young MA, Corrigan PW. DSM-V and the stigma of mental illness. J Ment Health 2010 Aug;19(4):318-327. [doi: 10.3109/09638237.2010.492484] [Medline: 20636112]

9. Corrigan PW, Druss BG, Perlick DA. The impact of mental illness stigma on seeking and participating in mental health care. Psychol Sci Public Interest 2014 Oct;15(2):37-70. [doi: 10.1177/1529100614531398] [Medline: 26171956]
10. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
11. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343 [FREE Full text] [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
12. Ben-Zeev D. Mobile health for all: public-private partnerships can create a new mental health landscape. JMIR Ment Health 2016 Jun 06;3(2):e26 [FREE Full text] [doi: 10.2196/mental.5843] [Medline: 27268283]
13. Aschbrenner KA, Naslund JA, Grinley T, Bienvenida JC, Bartels SJ, Brunette M. A survey of online and mobile technology use at peer support agencies. Psychiatr Q 2018 Sep;89(3):539-548 [FREE Full text] [doi: 10.1007/s11126-017-9561-4] [Medline: 29302771]
14. Brunette MF, Achtyes E, Pratt S, Stilwell K, Opperman M, Guarino S, et al. Use of smartphones, computers and social media among people with SMI: opportunity for intervention. Community Ment Health J 2019 Aug;55(6):973-978 [FREE Full text] [doi: 10.1007/s10597-019-00431-7] [Medline: 31175518]
15. Young AS, Cohen AN, Niv N, Nowlin-Finch N, Oberman RS, Olmos-Ochoa TT, et al. Mobile phone and smartphone use by people with serious mental illness. Psychiatr Serv 2020 Mar 01;71(3):280-283 [FREE Full text] [doi: 10.1176/appi.ps.201900203] [Medline: 31744429]
16. Noel VA, Acquilano SC, Carpenter-Song E, Drake RE. Use of mobile and computer devices to support recovery in people with serious mental illness: survey study. JMIR Ment Health 2019 Feb 20;6(2):e12255 [FREE Full text] [doi: 10.2196/12255] [Medline: 30785401]
17. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
18. Camacho E, Levin L, Torous J. Smartphone apps to support coordinated specialty care for prodromal and early course schizophrenia disorders: systematic review. J Med Internet Res 2019 Nov 12;21(11):e16393 [FREE Full text] [doi: 10.2196/16393] [Medline: 31714250]
19. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
20. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010 Jan;19(1-2):227-233. [doi: 10.1111/j.1365-2702.2009.03041.x] [Medline: 20500260]
21. Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. PLoS Med 2010 Nov 09;7(11):e1000368 [FREE Full text] [doi: 10.1371/journal.pmed.1000368] [Medline: 21085696]
22. Kreyenbuhl J, Nossel IR, Dixon LB. Disengagement from mental health treatment among individuals with schizophrenia and strategies for facilitating connections to care: a review of the literature. Schizophr Bull 2009 Jul;35(4):696-703 [FREE Full text] [doi: 10.1093/schbul/sbp046] [Medline: 19491314]
23. Smith TE, Easter A, Pollock M, Pope LG, Wisdom JP. Disengagement from care: perspectives of individuals with serious mental illness and of service providers. Psychiatr Serv 2013 Aug 01;64(8):770-775. [doi: 10.1176/appi.ps.201200394] [Medline: 23632442]
24. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online patient recruitment in clinical trials: systematic review and meta-analysis. J Med Internet Res 2020 Nov 04;22(11):e22179 [FREE Full text] [doi: 10.2196/22179] [Medline: 33146627]
25. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013 Feb 05;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]
26. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (The SpringboarD Trial): randomized controlled trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
27. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
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. 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]

30. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
31. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
32. 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]
33. Van Lieshout RJ, Goldberg JO. Quantifying self-reports of auditory verbal hallucinations in persons with psychosis. Can J Behav Sci/ Revue canadienne des sciences du comportement 2007;39(1):73-77. [doi: 10.1037/cjbs2007006]
34. Green CE, Freeman D, Kuipers E, Bebbington P, Fowler D, Dunn G, et al. Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales (GPTS). Psychol Med 2008 Jan;38(1):101-111. [doi: 10.1017/S0033291707001638] [Medline: 17903336]
35. Salzer MS, Brusilovskiy E. Advancing recovery science: reliability and validity properties of the Recovery Assessment Scale. Psychiatr Serv 2014 Apr 01;65(4):442-453. [doi: 10.1176/appi.ps.201300089] [Medline: 24487405]
36. Corrigan PW, Salzer M, Ralph RO, Sangster Y, Keck L. Examining the factor structure of the recovery assessment scale. Schizophr Bull 2004;30(4):1035-1041. [doi: 10.1093/oxfordjournals.schbul.a007118] [Medline: 15957202]
37. Rosenberg M. Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press; 2015.
38. Hawthorne G, Griffith P. The Friendship Scale: Development and Properties. Melbourne: Centre for Program Evaluation; 2000.
39. Sheehan DV. The Anxiety Disease. New York: Scribner; 1983.
40. Lewis J. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Human Comput Interac 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
41. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
42. Lund A. Measuring usability with the USE questionnaire. Usablity Interf 2001;8(2):3-6. [doi: 10.1177/1078087402250360]
43. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
44. Aboody D, Siev J, Doron G. Building resilience to body image triggers using brief cognitive training on a mobile application: a randomized controlled trial. Behav Res Ther 2020 Nov;134:103723. [doi: 10.1016/j.brat.2020.103723] [Medline: 32920164]
45. Cerea S, Ghisi M, Bottesi G, Carraro E, Broggio D, Doron G. Reaching reliable change using short, daily, cognitive training exercises delivered on a mobile application: the case of Relationship Obsessive Compulsive Disorder (ROCD) symptoms and cognitions in a subclinical cohort. J Affect Disord 2020 Nov 01;276:775-787. [doi: 10.1016/j.jad.2020.07.043] [Medline: 32738662]
46. Roncero M, Belloch A, Doron G. Can brief, daily training using a mobile app help change maladaptive beliefs? Crossover randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 13;7(2):e11443 [FREE Full text] [doi: 10.2196/11443] [Medline: 30758294]
47. Giraldo-O'Meara M, Doron G. Can self-esteem be improved using short daily training on mobile applications? Examining real world data of GG Self-esteem users. Clin Psychol 2021 May 17:131-139. [doi: 10.1111/cp.12219]
48. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
49. Cheema JR. Some general guidelines for choosing missing data handling methods in educational research. J Mod App Stat Meth 2014 Nov 01;13(2):53-75. [doi: 10.22237/jmasm/1414814520]
50. Vallejo G, Fernández MP, Livacic-Rojas PE, Tuero-Herrero E. Comparison of modern methods for analyzing repeated measures data with missing values. Multivariate Behav Res 2011 Nov 30;46(6):900-937. [doi: 10.1080/00273171.2011.625320] [Medline: 26736117]
51. Little RJ. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc 1993 Mar 20;88(421):125-134. [doi: 10.2307/2290705]
52. Crosier BS, Brian RM, Ben-Zeev D. Using Facebook to reach people who experience auditory hallucinations. J Med Internet Res 2016 Jun 14;18(6):e160 [FREE Full text] [doi: 10.2196/jmir.5420] [Medline: 27302017]
53. Ben-Zeev D, Buck B, Chander A, Brian R, Wang W, Atkins D, et al. Mobile RDoC: using smartphones to understand the relationship between auditory verbal hallucinations and need for care. Schizophr Bull Open 2020 Nov 9;1(1):sgaa060 [FREE Full text] [doi: 10.1093/schizbullopen/sgaa060] [Medline: 33937774]
54. Ben-Zeev D, Atkins DC. Bringing digital mental health to where it is needed most. Nat Hum Behav 2017 Dec;1(12):849-851. [doi: 10.1038/s41562-017-0232-0] [Medline: 31024173]
55. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]

56. Depp CA, Moore RC, Perivoliotis D, Granholm E. Technology to assess and support self-management in serious mental illness. Dialogues Clin Neurosci 2016 Jun;18(2):171-183 [FREE Full text] [Medline: 27489457]
57. Noel VA, Carpenter-Song E, Acquilano SC, Torous J, Drake RE. The technology specialist: a 21st century support role in clinical care. NPJ Digit Med 2019 Jun 26;2:61 [FREE Full text] [doi: 10.1038/s41746-019-0137-6] [Medline: 31388565]
58. Jonathan GK, Pivaral L, Ben-Zeev D. Augmenting mHealth with human support: notes from community care of people with serious mental illnesses. Psychiatr Rehabil J 2017 Sep;40(3):336-338 [FREE Full text] [doi: 10.1037/prj0000275] [Medline: 28891660]
59. Ben-Zeev D, Buck B, Chu PV, Razzano L, Pashka N, Hallgren KA. Transdiagnostic mobile health: smartphone intervention reduces depressive symptoms in people with mood and psychotic disorders. JMIR Ment Health 2019 Apr 12;6(4):e13202 [FREE Full text] [doi: 10.2196/13202] [Medline: 30977736]
60. Brewin CR. Understanding cognitive behaviour therapy: a retrieval competition account. Behav Res Ther 2006 Jun;44(6):765-784. [doi: 10.1016/j.brat.2006.02.005] [Medline: 16620779]
61. 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]
62. Mohr DC, Azocar F, Bertagnolli A, Choudhury T, Chrisp P, Frank R, Banbury Forum on Digital Mental Health. Banbury forum consensus statement on the path forward for digital mental health treatment. Psychiatr Serv 2021 Jun;72(6):677-683. [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
63. Opportunities and challenges of developing information technologies on behavioral and social science clinical research. National Institute of Mental Health. 2018. URL: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/bsscr-ops-dd-final_155109.pdf [accessed 2021-10-21]
64. Ben-Zeev D. The digital mental health genie is out of the bottle. Psychiatr Serv 2020 Dec 01;71(12):1212-1213. [doi: 10.1176/appi.ps.202000306] [Medline: 32576123]
65. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
66. Enforcement policy for digital health devices for treating psychiatric disorders during the coronavirus disease 2019 (COVID-19) public health emergency. U.S Food & Drug Administration. 2020. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enforcement-policy-digital-health-devices-treating-psychiatric-disorders-during-coronavirus-disease [accessed 2021-10-20]

Abbreviations

ANOVA: analysis of variance
BDI-II: Beck Depression Inventory-Second Edition
GAD-7: Generalized Anxiety Disorder–7
GPTS: Green Paranoid Thought Scale
HPSVQ: Hamilton Program for Schizophrenia Voices
mHealth: mobile health
RAS: Recovery Assessment Scale
RSES: Rosenberg Self-Esteem Scale
SDS: Sheehan Disability Scale
SMI: Serious Mental Illness

Edited by R Kukafka, G Eysenbach; submitted 30.03.21; peer-reviewed by A Bourla, I Bell, Z Ali, M Tænning; comments to author 10.05.21; revised version received 19.05.21; accepted 03.10.21; published 12.11.21.

Please cite as:
Ben-Zeev D, Chander A, Tauscher J, Buck B, Nepal S, Campbell A, Doron G
A Smartphone Intervention for People With Serious Mental Illness: Fully Remote Randomized Controlled Trial of CORE
J Med Internet Res 2021;23(11):e29201
URL: https://www.jmir.org/2021/11/e29201
doi:10.2196/29201
PMID:34766913

Review

Efficacy of Interventions That Incorporate Mobile Apps in Facilitating Weight Loss and Health Behavior Change in the Asian Population: Systematic Review and Meta-analysis

Siew Min Ang[1*], BSc, MND, APD; Juliana Chen[2*], BSc, MND, APD, PhD; Jia Huan Liew[3], PhD; Jolyn Johal[4], BND, APD; Yock Young Dan[5], MD; Margaret Allman-Farinelli[2], PhD, APD; Su Lin Lim[1], PhD

[1]Department of Dietetics, National University Hospital, Singapore, Singapore

[2]School of Life and Environmental Sciences, Charles Perkins Centre, The University of Sydney, Sydney, Australia

[3]Science Unit, Lingnan University, Hong Kong, Hong Kong

[4]Joanna Briggs Institute, University of Adelaide, Adelaide, Australia

[5]Department of Medicine, National University Hospital, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Siew Min Ang, BSc, MND, APD
Department of Dietetics
National University Hospital
5 Lower Kent Ridge Road
Singapore, 119074
Singapore
Phone: 65 67725166
Email: siew_min_ang@nuhs.edu.sg

Abstract

Background: Smartphone apps have shown potential in enhancing weight management in Western populations in the short to medium term. With a rapidly growing obesity burden in Asian populations, researchers are turning to apps as a service delivery platform to reach a larger target audience to efficiently address the problem.

Objective: This systematic review and meta-analysis aims to determine the efficacy of interventions that incorporate apps in facilitating weight loss and health behavior change in the Asian population.

Methods: A total of 6 databases were searched in June 2020. The eligible studies included controlled trials in which an app was used in the intervention. The participants were aged 18 years or older and were of Asian ethnicity. A meta-analysis to test intervention efficacy, subgroup analyses, and post hoc analyses was conducted to determine the effects of adding an app to usual care and study duration. The primary outcome was absolute or percentage weight change, whereas the secondary outcomes were changes to lifestyle behaviors.

Results: A total of 21 studies were included in this review, and 17 (81%) were selected for the meta-analysis. The pooled effect size across 82% (14/17) of the randomized controlled trials for weight change was small to moderate (Hedges *g*=–0.26; 95% CI –0.41 to –0.11), indicating slightly greater weight loss achieved in the intervention group; however, this may not be representative of long-term studies (lasting for more than a year). Supplementing multicomponent usual care with an app led to greater weight loss (Hedges *g*=–0.28; 95% CI –0.47 to –0.09). Asian apps were largely culturally adapted and multifunctional, with the most common app features being communication with health professionals and self-monitoring of behaviors and outcomes.

Conclusions: More evidence is required to determine the efficacy of apps in the long term and address the low uptake of apps to maximize the potential of the intervention. Future research should determine the efficacy of each component of the multicomponent intervention to facilitate the designing of studies that are most effective and cost-efficient for weight management.

Trial Registration: PROSPERO CRD42020165240; https://tinyurl.com/2db4tvn6

(J Med Internet Res 2021;23(11):e28185) doi:10.2196/28185

KEYWORDS

systematic review; meta-analysis; mobile app; obesity; weight loss; Asian; diet; physical activity; adults; mobile phone

Introduction

Background

Asian countries typically have lower rates of obesity than Western countries [1,2]. However, globalization has contributed to rapid increases in Asian obesity rates over the last 10-15 years such that between 20% and 35% of Asian adults are overweight or obese [2,3]. Higher body fat percentage, prominent central adiposity, and possible genetic factors predispose Asians to insulin resistance, type 2 diabetes, and cardiovascular diseases that further aggravate the health care burden [4,5], with direct health care costs estimated to be US $100 billion in Asian countries alone [3]. As part of the global action plan for the prevention and control of noncommunicable diseases, the World Health Organization has recommended a focus on improving lifestyle behaviors, including adopting a healthy diet and increasing physical activity to modify the risk factors for obesity and noncommunicable diseases [6].

With the exponential growth of mobile technology in the past decade, researchers have explored the potential of digital health interventions using mobile apps as a service delivery platform to reach a larger target audience [7,8]. This is particularly promising in Asian countries where smartphone adoption is estimated to reach 84% by 2025 [9,10]. The ubiquity, accessibility, multifunctionality, and scalability of apps for health intervention provide health care professionals and researchers an unprecedented avenue for treating, monitoring, and interacting with patients en masse remotely [11,12]. With technological advancement, the number of health- and fitness-related apps targeted at behavioral change has burgeoned, with at least 325,000 apps available on the commercial market in 2017 [13].

To date, several systematic reviews on smartphone efficacy to improve weight and health have concluded that interventions that incorporate apps show potential in weight management as well as in improving diet, physical activity, and chronic disease outcomes and are acceptable in the short to medium term [7,8,14-17]. However, most of the studies included in these systematic reviews have been focused on Western populations. Given the differences in genetics, culture, lifestyle, health beliefs, and health-seeking behaviors between Asian and White populations [5,18,19], it is important to assess if interventions that incorporate apps are efficacious in achieving weight loss in Asians before considering them as part of a national strategy to combat obesity.

Objective

The aim of this review is to systematically synthesize evidence to address this gap in the literature and provide recommendations for future studies. The primary outcome of this review was absolute or percentage weight loss or other surrogate measures of body fat composition such as BMI or waist circumference. The secondary outcomes included dietary intake, physical activity, self-efficacy with regard to implementing healthy lifestyle behaviors, and user engagement with apps.

Methods

Literature Search

The review protocol was prospectively registered (PROSPERO ID: CRD42020165240), with modifications made over the course of the review (Multimedia Appendix 1), and conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 2) [20]. Systematic searches were conducted in June 2020 across 6 databases: MEDLINE, CINAHL, Embase, PsycINFO, Global Health, and the Cochrane Central Register of Controlled Trials. The search strategy incorporated Medical Subject Headings, keywords, and free-text search terms. The search terms included *app**, *application**, *mobile app**, *smartphone*, *mHealth*, *weight loss*, *weight change*, *body mass index change*, and *Asian**. A sample search for MEDLINE is detailed in Multimedia Appendix 3. *BMC Proceedings*, ProQuest Dissertations & Theses, and Google Scholar were searched for conference proceedings, dissertations, and any unpublished gray literature, whereas the ISRCTN registry, ClinicalTrials.gov, and the World Health Organization International Clinical Trials Registry Platform were queried for eligible clinical trials and research. Reference lists of the eligible studies and review articles were also manually searched for additional papers that warranted inclusion. In addition, a filtered search of the *Journal of Medical Internet Research* and *JMIR mHealth and uHealth* was conducted to locate papers that were published before the respective journals were indexed in MEDLINE.

Inclusion Criteria and Study Selection

Studies were included if they were randomized controlled trials (RCTs), quasi-randomized trials, or nonrandomized controlled trials (non-RCTs). Interventions with no control group, before-and-after interventions, and observational studies (cohort, case-control, cross-sectional, and ecological) were excluded. To be included in this review, studies needed to use a mobile app either in a single-component (ie, standalone use of apps) or multicomponent (ie, apps as part of an intervention with other components, eg, face-to-face consultation, phone calls, or email reviews) intervention. Given that app engagement typically declines rapidly by the second month [21], this was chosen as the minimum study duration to ensure that the intervention effects of app use and longer-term outcomes could be assessed. Participants had to be aged 18 years or older and of Asian ethnicity. Studies were excluded if participants were reported to have eating disorders or mental health conditions, bariatric surgery, or were within the pregnancy or postpartum period. This review was limited to research published in the English language and from 2008 to date because 2008 was the year in which smartphone apps emerged [22].

Data Extraction

All titles and abstracts of the retrieved records were independently screened by 2 reviewers (SMA and JC) to identify the records that potentially met the inclusion criteria. Relevant full-text articles were retrieved and independently assessed by both reviewers using the complete inclusion and exclusion criteria. Both reviewers independently extracted data from the articles based on a standardized data extraction form, including

study characteristics (author, year, country, study design, study aims, sample number, attrition rate, disease group, conflict of interest, and funding), intervention characteristics (intervention type, duration, app type, app features, and cultural adaptations), and predefined outcomes. The level of agreement between the reviewers for the main stages of screening were assessed using the Cohen κ coefficient. Discrepancies were discussed and resolved between the reviewers. Any missing data or further information required was requested by email from the corresponding authors, with a follow-up reminder sent after 2 months.

Study Quality Assessment

The Cochrane Collaboration Risk-of-Bias Tool [23] and the Risk of Bias in Non-Randomized Studies of Interventions Tool [24] were implemented independently by 2 reviewers (SMA and JC) to assess the risk of bias in the RCTs and non-RCTs, respectively. Discrepancies were discussed and resolved through a third reviewer (JJ). Each domain of the RCTs received an evaluation of low, moderate, or high risk, whereas the non-RCTs were judged as having low, moderate, serious, or critical risk.

Outcomes

The primary outcome of this review was absolute (kg) or percentage weight change. Other surrogate measures of body weight, such as BMI (kg/m^2) and waist circumference (cm), were also included. The secondary outcomes included app use and changes to lifestyle behaviors, including diet, physical activity, and self-efficacy for implementing healthy behaviors.

Data Analysis

The effect sizes used in the meta-analysis were Hedges *g* values calculated from the mean differences in outcomes (ie, changes in absolute or percentage weight, BMI, or waist circumference before and after treatment) between the treatment arms. Separate analyses were conducted for the RCTs and non-RCTs; studies without a standardized mean and SD were excluded from the meta-analysis. A unique study identifier was assigned to each intervention-control pair included in the meta-analysis (Multimedia Appendix 4 [25-41]).

Random effects models, which control for heterogeneity between studies, were used to fit the Hedges *g* scores. To account for within-study dependencies, comparisons were made separately for each outcome, namely, changes in absolute or percentage weight, BMI, and waist circumference. For studies with multiple time points reported, only the final outcome within the active intervention was included to avoid pseudoreplication. Furthermore, the overall data were also divided into 2 subsets comprising single-component (ie, standalone use of apps) and multicomponent (ie, apps as a part of an intervention with other components) studies, respectively, and analyzed separately. Subsequently, subgroup analyses were conducted to analyze the effects of adding an app to usual care (intervention group) compared with usual care alone (control group).

The possible moderating effect of study duration was also tested using moderation analysis, after which the possible differences in intervention outcomes in studies conducted over 3 months or less versus studies lasting for longer than 3 months were assessed post hoc. These were done as a preliminary assessment of the importance of app engagement levels on study outcomes because adherence to app use generally tails off after 2 to 3 months [8,11,42,43]. The post hoc analyses were conducted in lieu of a more robust meta-regression approach given the dearth of quantifiable app engagement data.

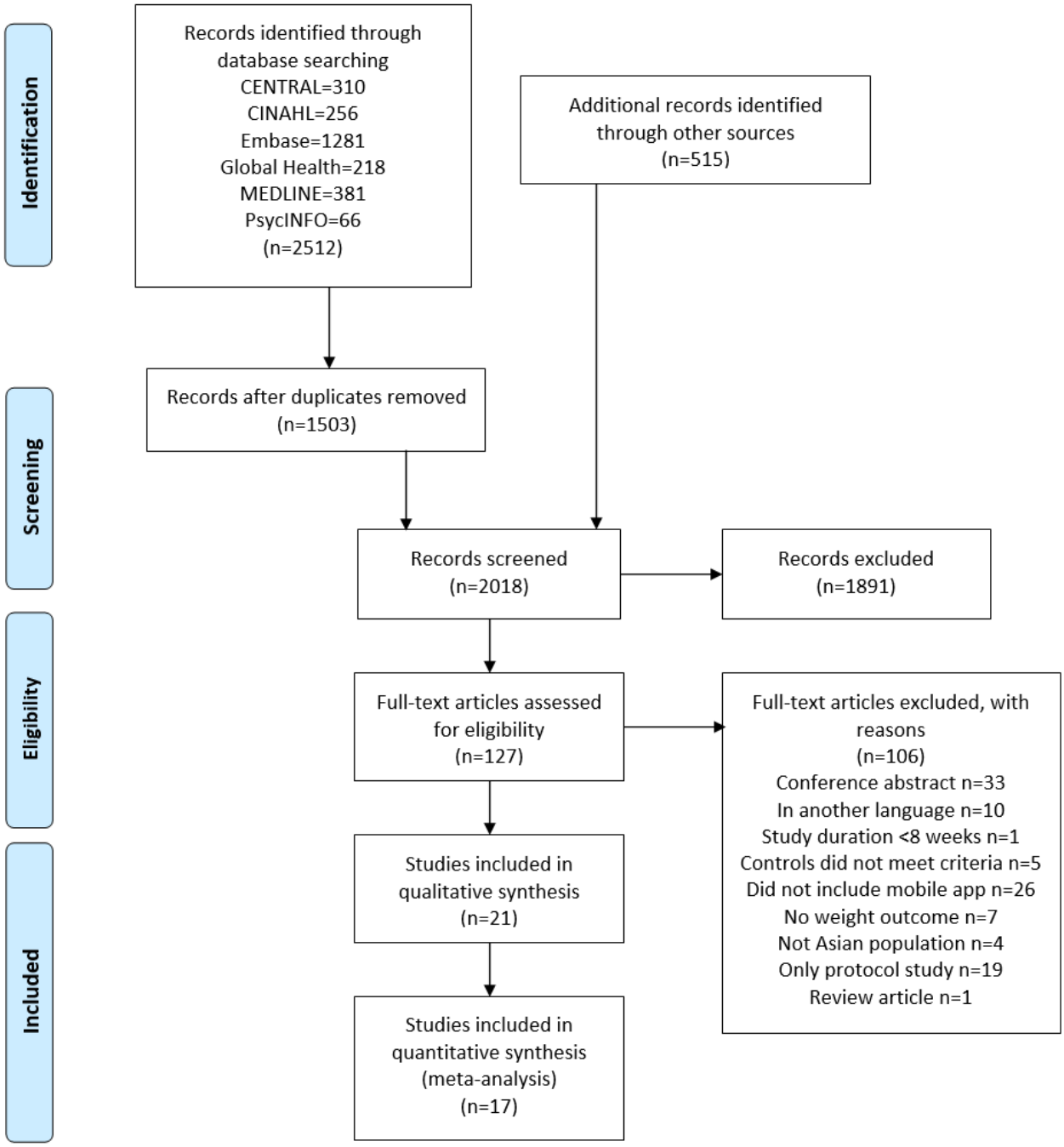
Heterogeneity among the studies for each comparison was assessed using the I^2 statistic, with values of 30% to 60%, 50% to 90%, and 75% to 100% considered to indicate moderate, substantial, and considerable levels of heterogeneity, respectively [44]. Publication bias was also assessed using a funnel plot. All analyses were conducted using the *metafor* package (2.4.0) in the R statistical environment (R Foundation for Statistical Computing) [45,46].

Results

Study and Sample Characteristics

A total of 3027 electronic records were identified through the search strategy and, after removal of duplicates, 2018 (66.67%) titles were screened. From these 2018 records, 127 (6.29%) full-text articles were retrieved (Figure 1). Of the 127 papers, 21 (16.5%) met all inclusion criteria for the systematic literature review [25-41,47-50]. Of these 21 papers, 17 (81%) were included in the meta-analysis [25-41]. The Cohen κ coefficients for the initial screening stage of titles and abstracts and the full-text screening stage were 0.72 (substantial agreement) and 0.92 (almost perfect agreement), respectively, with 97% level of agreement for each screening stage.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram. CENTRAL: Cochrane Central Register of Controlled Trials. CINAHL: Cumulative Index to Nursing and Allied Health Literature.



Intervention characteristics and descriptions of the studies included in this review are summarized in Table 1 and Table 2, respectively. Of the 21 papers, 17 (81%) were RCTs and 4 (19%) were non-RCTs, with study locations in China [25,37-39,48], Hong Kong [34], India [26,31], Japan [35], Singapore [30], South Korea [27-29,32,33,36,40,50], Taiwan [41,49], and the United States [47]. The total number of participants across the 21 studies was 21,173 (RCTs: 4090 and non-RCTs: 17,083), with a mean age of 45.9 (SD 9.84, range 25.8-60.5) years; on average, 45.1% (SD 20.1) of the participants were women. The mean BMI of the population was 27.1 (SD 2.47, range 23.0-30.5) kg/m^2. Of the 21 studies, 9 (43%) were conducted among patients who were overweight or obese [28,30-35,47,49], 6 (29%) involved mainly patients with diabetes mellitus—4 (67%) [27,36,40,48] on type 2, whereas 2 (33%) [37,38] included a mix of type 1 and 2—and 3 (14%) included participants from the general population [26,39,41], whereas the remaining 3 (14%) individually targeted patients with metabolic abnormalities [50], coronary heart disease [25], or colorectal polyps [29].

Table 1. Characteristics of the interventions incorporating apps included in the review (N=21).

Author (year), country, ethnicity	Study characteristics, sample size (included in analysis)	Participant characteristics	App characteristic (name), cultural adaptation within app	Measured outcomes	Attrition rate (%)
Bender et al [47] (2018), United States, Filipino American	Pilot RCT[a], Filipino Americans who were overweight or obese and aged ≥18 years at risk of T2DM[b] or prediabetes, 3 months, n=67	Mean age, years (SD): 41.7 (12.0), mean BMI, kg/m^2 (SD): 30.5 (4.4), women (%): 52.2	Multicomponent, commercial (Fitbit), English, no	Weight, BMI, waist circumference, FPG[c], and HbA_{1c}[d] level	5; I[e]: 6; C[f]: 3
Dong et al [48] (2018), China, Chinese	RCT, patients with T2DM aged 18-60 years, 12 months, n=120 (119)	Mean age, years (SD): 42.7 (6.7), BMI ≥25 kg/m^2: I (%): 48.3, C (%): 42.4, women (%): 47.9	Multicomponent, commercial (WeChat), Chinese, yes	FPG, 2-hour PG[g], HbA_{1c} level, total self-efficacy score, diet score, exercise score, medication-taking score, blood glucose–monitoring score, foot-care score, and smoking score	0
Dorje et al [25] (2019), China, Chinese	RCT, patients with coronary heart disease aged ≥18 years, 2 months intervention+4 months step-down phase, n=312	Mean age, years (SD): 60.5 (9.2), mean BMI, kg/m^2 (SD): I: 25.5 (3.0), C: 25.4 (3.5), women (%): 19	Multicomponent, commercial (WeChat), Chinese, yes	Weight, BMI, waist-to-hip ratio, 6-minute walk test, knowledge and awareness of coronary heart disease, PG, lipids, psychosocial well-being, quality of life, smoking status, dietary habits, and physical activity	15; I: 14; C: 16
Kaur et al [26] (2020), India, Indian	Cluster RCT, adults aged 35-70 years, 6 months, n=732	Mean age, years (SD not reported): 52.7, mean BMI, kg/m^2 (SD): I: 27.03 (4.2), C: 27.45 (4.8), women (%): 76.1	Multicomponent, commercial (WhatsApp), English, Hindi, and Punjabi, emails in English, yes	Weight, BMI, dietary intake changes, ASE[h] score, BP[i], FPG, and lipids	9; I: 8; C: 10
Kim et al [27] (2019), South Korea, Korean	Multicenter RCT, stable patients with T2DM aged 19-80 years with HbA_{1c} level between 7% and 10%, 24 weeks, n=191 (172)	Mean age, years (SD): I: 60.0 (8.4), C: 56.7 (9.1), mean BMI, kg/m^2 (SD): I: 25.5 (3.2), C: 25.8 (4.1), women (%): I: 44, C: 52	Multicomponent, researcher design (mDiabetes), Korean, yes	Weight, body composition, score of the summary of diabetes self-care activities, World Health Organization quality-of-life scale, HbA_{1c} level, lipids, and BP	21
Lee et al [28] (2018), South Korea, Korean	Multicenter pilot RCT, university medical school students who were overweight or obese with metabolic syndrome, 24 weeks, n=422 (324)	Age: ≥20 years, mean BMI, kg/m^2 (SD): inactive I: 28.8 (2.72), C: 29.1 (3.10), moderately active I: 29.8 (4.39), C: 29.0 (2.46), health-enhancing physically active I: 29.3 (2.90), C: 29.8 (7.12), women (%): I: 52, C: 48	Multicomponent, researcher design (SmartCare), Korean, app is in English, yes	Weight, BMI, waist circumference, body composition, BP, FPG, HbA_{1c} level, and lipids	23; I: 17; C: 30
Lee et al [29] (2019), South Korea, Korean	Pilot RCT, patients aged 20-65 years with colorectal polyps diagnosis within the last 2 years of the study, 3 months, n=65	Mean age, years (SD): I: 49.1 (8.3), C: 50.7 (8.1), mean BMI, kg/m^2 (SD): I: 26.9 (3.4), C: 24.5 (3.9), women (%): I: 34.4, C: 45.5	Multicomponent, commercial (Noom), English or Korean (NFS[j]), no	Weight, changes in dietary intake through food frequency questionnaire, and Godin leisure-time exercise questionnaire	3; I: 0; C: 6
Lim et al [30] (2020), Singapore, multiracial	RCT, patients who were overweight or obese with nonalcoholic fatty liver disease and were aged 21-70 years, 6 months, n=108	Mean age, years (SD): I: 46.8 (11.1), C: 46.2 (10.1), mean BMI, kg/m^2 (SD): I: 30.1 (4.0), C: 30.8 (4.8), women (%): I: 42, C: 32	Multicomponent, researcher design (nBuddy), English, yes	Weight, BMI, waist circumference, BP, and liver enzymes	6; I: 9; C: 4
Muralidharan et al [31] (2019), India, Indian	Multicenter RCT, adults who were overweight or obese aged 20-65 years with prediabetes, 12 weeks, n=741 (561)	Mean age, years (SD): I: 37.8 (9.2), C: 37.8 (9.6), mean BMI not reported, only baseline weight was reported, women (%): I: 43.9, C: 42.1	Multicomponent, researcher design (mDiab), English, yes	Weight, target 5% weight loss	24; I: 28; C: 21

Author (year), country, ethnicity	Study characteristics, sample size (included in analysis)	Participant characteristics	App characteristic (name), cultural adaptation within app	Measured outcomes	Attrition rate (%)
Oh et al [32] (2015), South Korea, Korean	Multicenter RCT, adults who were obese aged 20-70 years with BMI≥25 kg/m^2 and metabolic syndrome diagnosis, 24 weeks, n=422 (334)	Mean age, years (SD): I: 46.78 (13.11), C: 50.35 (14.24), mean BMI, kg/m^2 (SD): I: 29.42 (3.53), C: 29.40 (3.39), women (%): I: 46.7, C: 51.4	Multicomponent, researcher design (SmartCare), Korean, app is in English, yes	Weight, BMI, waist circumference, body composition, change in diet habit, change in physical activity (IPAQ^k), and patient satisfaction	21; I: 15; C: 27
Shin et al [33] (2017), South Korea, Korean	Pilot RCT, men who were overweight or obese aged 19-45 years with BMI≥27 kg/m^2, 12 weeks, n=105 (98)	Mean age, years (SD): 27.8 (5.0), mean BMI, kg/m^2 (SD): 29.8 (2.7), women: 0	Multicomponent, researcher design (Fit.Life), Korean, app is in English, yes	Weight, BMI, body composition, physical activity changes (IPAQ), calorie intake changes, BP, FPG, lipids, and liver enzymes	7; C: 9; I1: 3; I2: 9
Suen et al [34] (2019), Hong Kong, Chinese	Feasibility RCT, healthy adults who were overweight or obese aged ≥18 years with BMI≥25 kg/m^2 without ear injuries, 8 weeks, n=59	Mean age, years (SD): 49.15 (10.54), mean BMI, kg/m^2 (SD): 30.35 (4.53), women (%): 85	Multicomponent, researcher Design (Auricular Acupressure for Weight Reduction version 1), Chinese, yes	Weight, BMI, body composition, waist and hip circumference, blood leptin and adiponectin, fullness rating, and patient satisfaction	10; C: 5; I: 11
Tanaka et al [35] (2018), Japan, Japanese	RCT, adults who were overweight, obese, or abdominally obese aged 20-64 years with cardiometabolic risk factors or metabolic syndrome, 8 weeks intervention+4 weeks postintervention follow-up, n=112	Mean age, years (SD): I: 45.6 (10.2), C: 47.8 (9.3), mean BMI, kg/m^2 (SD): I: 28.0 (3.3), C: 28.2 (3.0), women (%): 0.9	Single-component, commercial (FiNC), Japanese yes	Weight, waist circumference, BP, lipids, HbA_{1c} level, and obesogenic eating behaviors	28; I: 32; C:18
Yang et al [49] (2017), Taiwan, Chinese	Crossover RCT, patients who were overweight or obese with BMI≥24 kg/m^2 and metabolic abnormalities, 3 months intervention+crossover 3 months usual care, n=53 (46)	Mean age, years (SD): 33.2 (9.6), mean BMI, kg/m^2 (SD): I: 27.2 (3.4), C: 30.3 (4.9), women (%): I: 61.5, C: 59.2	Multicomponent, researcher design (self-monitoring app), NFS (Line, social communication app), NFS on app, Chinese and English noted on website, yes	Weight, BMI, waist circumference, change in physical activity, FPG, BP, and lipids	13; I: 19; C: 7
Yang et al [36] (2020), South Korea, Korean	Cluster RCT, adults aged ≥18 years with T2DM for ≥1 year and HbA_{1c} level 7%-10%, 12 weeks, n=247 (239)	Mean age, years (SD): I: 54.1 (10.1), C: 60.6 (10.2), mean BMI, kg/m^2 (SD): I: 26.3 (3.7), C: 25.7 (3.9), women (%): I: 46.6, C: 53.6	Multicomponent, researcher design (HiCare smart K), Korean, yes	Weight (not mandatory), BMI, waist circumference, BP, lipids, HbA_{1c} level, FPG, Diabetes Treatment Satisfaction Questionnaire, and medication adherence scale	3; I: 3; C: 3
Zhang et al [37] (2019), China, Chinese	RCT, adults aged 18-65 years, diagnosed with diabetes for more than 6 months and with HbA_{1c} level ≥8%, 6 months, n=234 (194)	Mean age, years (SD): 53 (11), mean BMI, kg/m^2 (SD): 25.03 (3.36), women (%): 38	Multicomponent, commercial (Welltang), Chinese, yes	Weight, BMI, BP, waist circumference, FPG, HbA_{1c} level, lipids, and liver enzymes	17; C: 19; I1: 14; I2: 18
Zhou et al [38] (2016), China, Chinese	Pilot RCT, adults aged 18-74 years with diagnosed diabetes without severe complications, 3 months, n=100	Mean age, years (SD): I: 55.0 (13.1), C: 53.5 (12.4), mean BMI, kg/m^2 (SD): I: 23.04 (4.09), C: 23.01 (4.04), women (%): I: 46, C: 40	Multicomponent, commercial (Welltang), Chinese, yes	Weight, BMI, waist and hip circumference, diabetes knowledge and self-care behavior score, HbA_{1c} level, FPG, 2-hour PG, BP, and low-density lipoprotein-c	Not reported
He et al [39] (2017), China, Chinese	Cohort-based non-RCT, general population aged ≥18 years who were keen on weight loss, 6 months, n=15,818 (15,310)	Mean age, years (SD): I: 35.1 (8.5), C: 39.0 (9.5), mean BMI not reported, women (%): I: 66.5, C: 40.5	Single-component, commercial (WeChat), Chinese, yes	Weight and waist circumference	3; I: 4; C: 2

Author (year), country, ethnicity	Study characteristics, sample size (included in analysis)	Participant characteristics	App characteristic (name), cultural adaptation within app	Measured outcomes	Attrition rate (%)
Kim et al [40] (2014), South Korea, Korean	Matched controlled non-RCT, adults aged 20-70 years with T2DM for more than 1 year with HbA_{1c} level 7%-10% at baseline, 3 months, n=73 (70)	Mean age, years (SD): I: 51.8 (10.3), C: 53.8 (9.0), mean BMI, kg/m^2 (SD): I: 25.0 (3.3), C: 24.9 (3.4), women (%): 49.2	Multicomponent, NFS, likely researcher design (Mobile SmartCare, version 1.0.7), Korean, yes	BMI, BP, HbA_{1c} level, lipids, and patient satisfaction	4; I: 8; C: 0
Kim et al [50] (2019), South Korea, Korean	Cohort-based non-RCT, control: individuals with metabolic abnormalities according to the Adult Treatment Panel III criteria, intervention: individuals recruited from among those who had previously completed a 24-week mobile service program as part of the First Year Public Health Center Mobile Healthcare pilot project, 24 weeks, n=1117	Mean age, years (SD): I: 44.68 (8.22), C: 44.69 (8.22), mean BMI, kg/m^2 (SD): I: 25.71 (3.34), C: 25.18 (3.48), women (%): I: 50.8, C: 65.5	Multicomponent, researcher design (Public Health Center mHealth app), Korean, yes	Health behavior scores, mini–dietary assessment scores, BP, FPG, triglycerides, high density lipoprotein-c, and waist circumference	Not reported
Wijaya and Widiantoro [41] (2018), Taiwan, Indonesian	Pretest-posttest design, Indonesian international students aged ≥20 years who owned a smartphone with internet access, not participating in other training program, and literate in English, 10 weeks, n=75 (70)	Mean age, years (SD): 25.86 (4.33), mean BMI, kg/m^2 (SD): 23.25 (3.05), women (%): 45.7	Multicomponent, researcher Design (iNCKU smartphone app) English, yes	Body weight, BMI, BP, physical fitness, physical activity measures (step count, distance covered, caloric expenditure, and time spent on activity), self-efficacy, social support, and outcome expectation	7; I: 8; C: 5

[a]RCT: randomized controlled trial.

[b]T2DM: type 2 diabetes mellitus.

[c]FPG: fasting plasma glucose.

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

[e]I: intervention.

[f]C: control.

[g]PG: plasma glucose.

[h]ASE: attitude, social influence, and self-efficacy.

[i]BP: blood pressure.

[j]NFS: not further specified.

[k]IPAQ: International Physical Activity Questionnaire.

Table 2. Description and primary outcome of the interventions included in the review (N=21).

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Bender et al [47] (2018), United States, Filipino American, pilot RCT[a]	Fit&Trim, a DPP[b]-based, culturally adapted, mobile phone–based weight loss lifestyle intervention delivered by Filipino research staff, augmented with tracker Fitbit Zip and private Facebook virtual support group. Provided individual tailored goals for weight, diet, physical activity or steps, and encouraged to monitor lifestyle habits and progress on app	Baseline+5 in-person Fit&Trim education intervention office visits	Active waitlist. Received Fitbit Zip physical activity tracker with 2 education sessions on hepatitis A and B	Weight change (kg) calculated: I[c]: –3.39, C[d]: –0.82, SD or further values not reported, achieved 5% weight loss: I: 36%, C: 6%, large effect of 0.93 (Cohen *d*) reported
Dong et al [48] (2018), China, Chinese, RCT	Received conventional health education with nursing care for diabetes. Patients received multimedia-type diabetes-related knowledge from nurses on app. Able to communicate with educators and friends on app. Phone reviews on app use and hospital physical examination offered	Baseline visit and 6-month and 12-month visits, nurses communicating with patients through WeChat	Conventional health education with nursing care for diabetes	There was no significant difference in BMI between groups at baseline, 6 months, and 12 months. No further values reported
Dorje et al [25] (2019), China, Chinese, RCT	Received usual care+SMART-CR/SP[e] program. During the intensive phase, participants received 4 educational cartoon modules per week through WeChat. In the step-down phase, participants received only 2 cartoon pictures with key motivational message per week; Cartoon education touched on cardiovascular health and disease, physical activity, healthy nutritional advice, support for medication adherence, psychological well-being, and modification risk factors; Individualized feedback, recommendations, and remote supervision were provided based on regular reviews of monitoring data. Coach support available on app for health and lifestyle advice. Additional alerts and WeChat messages were sent when measurements were outside target blood pressure or steps	Baseline, 2-month, and 6-month visits for measurements and assessments by blinded researchers. Remote supervision through messages, telemonitoring, feedback, and video calls from coaches as necessary based on regular reviews of data	Standard care provided by doctors with a brief education carried out by a nurse. Medication management and ad hoc review visits to a cardiologist or other health care providers according to the patient's self-assessment of their own cardiovascular health; WeChat used for sending review visit reminders, did not receive any form of health information or intervention	BMI change (kg/m^2), mean (SD), baseline: I: 25.5 (3.0), C: 25.4 (3.5), 2 months: I: 25.0 (2.9), C: 25.2 (3.2), between-groups *P*=.64, 6 months: I: 24.9 (3.5), C: 24.5 (3.2), between-groups *P*=.14; Waist circumference change (cm), mean (SD), baseline: I: 0.9 (0.1), C: 0.9 (0.1), 2 months: I: 0.9 (0.1), C: 0.9 (0.0), between-groups *P*=.36, 6 months: I: 0.9 (0.0), C: 0.9 (0.1), between-groups *P*=0.95
Kaur et al [26] (2020), India, Indian, cluster RCT	Provision of *SMART Eating* kit, which included a kitchen calendar, dining table mat, and measuring spoons. Received weekly review information through SMS, email, social networking app (WhatsApp), and *SMART Eating* website; Fortnightly addition of diet and health-related content to website, quizzes, and a web-based help assistant to ask questions	Single home visit to provide education and guide family champions on how to use the different components of the intervention; No further face-to-face interaction. Further advice and supervision was conducted through the website	Pictorial pamphlet on the dietary recommendations with information written in Hindi language. Asked to read the pamphlet in their own time, make changes to their diet accordingly, and convey the same information to their family members; Same educational content and materials were offered to intervention group	Weight change (kg), mean (95% CI; SD not reported), 6 months: I: –0.42 (–0.8 to –0.1) *P*=.01, C: 0.24 (–0.1 to 0.6) *P*=.14, net change between groups: –0.66 (–1.1 to –0.2), *P*=.01

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Kim et al [27] (2019), South Korea, Korean, multicenter RCT	Participants were divided into 4 groups based on antidiabetic treatment. Nil baseline education. Provided individualized targets for diet and physical activity at baseline. Encouraged to monitor blood glucose levels and lifestyle habits on app and informed that physicians can view and monitor their progress through telemonitoring. App provided immediate feedback according to an algorithm when a reading on blood glucose level, food intake, or activity was entered. Detailed information on diet and physical activity was made available through a range of educational and interactive component on app. Social networking service and bulletin board enabled users to share experiences and tips with each other, whereas research staff could answer questions from patients	Baseline, 2 in-person visits (week 12 and week 24), and 2 phone call reviews. Remote supervision and advice provided on app as well	Nil baseline education. Provided logbook to record blood glucose readings, Bluetooth glucometer, test strips, and a printed education booklet. Received 2 in-person visits at 12 weeks and 24 weeks	Weight change (kg), mean (SD), baseline to 6 months: I: 67.7 (11.8) to 67.1 (11.6), *P*=.005, C: 68.4 (13.0) to 68.0 (12.7), *P*=.04
Lee et al [28] (2018), South Korea, Korean, multicenter pilot RCT	Provided with a smartphone equipped with SmartCare app and Bluetooth-enabled bioimpedance analyzer. Instructed to monitor body composition, data were transmitted to the SmartCare system through the app; Health reports were automatically created based on the personal health information of participants according to the clinical decision support system algorithm function of the SmartCare system. Health managers provided prevention, consultation, and educational services remotely to participants based on health reports through messages and weekly emails. Monthly progress evaluation was offered along with in-person consultation with a physician at least once every 2 months (follow-up study by Oh et al [32], 2015)	Baseline, 2-monthly in-person visits with physician. Measurements at baseline, week 12, and week 24. Weekly remote supervision over app	Provided with weighing scale and pedometer and asked to record weight and physical activity steps progress in a diary. Offered 3 in-person visits at baseline, week 12, and week 24. No further visit details	Weight change (kg), mean (SD) 6 months, insufficiently active: C: –0.1 (1.94), *P*=.64, I: –1.6 (3.03), *P*<.001, between-groups *P*=.001, minimally active: C: –0.3 (2.24), *P*=.49, I: –2.5 (3.81), *P*=.001, between-groups *P*=.01, health-enhancing physical activity: C: –1.5 (3.12), *P*<.001, I: –2.6 (3.91), *P*<.001, between-groups *P*=.05
Lee et al [29] (2019), South Korea, Korean, pilot RCT	Received app and taught to use under supervision during first visit without further education. Health-related information, lifestyle recommendations, and feedback sent through app; Encouraged to track lifestyle habits on app, and users may mutually compete and share progress on the bulletin board. Monthly phone interview to assess the proper use of the app and provide motivation	In-person session to download and teach the app use during baseline visit. Measurement taken at baseline and 3 months; Nil further face-to-face interaction. Monthly phone calls (2 calls)	Received a diary to record food intake and exercise. Staff members provided health behavior change education at baseline visit; Health-related newsletters sent monthly, containing the same information about behavior as those received by the experimental group. Received monthly phone interview motivation and review	Weight change (kg), mean (SD) 3 months: I: –1.25 (1.14), *P*<.01, C: –0.42 (1.23), *P*<.07, between-groups *P*<.01; The effect was most pronounced in app users with good adherence (1.45 kg more weight reduction than the control group participants)

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Lim et al [30] (2020), Singapore, multiracial RCT	Guided on the use of the app and educated on dietary and physical activity modification through a 1-hour face-to-face session with the research dietitian at the first study visit. Set individualized weight and lifestyle goals on the app. Advised to monitor lifestyle and progress on app. Educational videos and daily tips on healthy lifestyle available on app. Reminders and push notifications in place for meal and weight logging. Remote coaching with dietitians on progress, received encouragement and advice on app. Provided weighing scale	1-hour individual face-to-face education session with the research dietitian at baseline visit; Remote dietitian coaching on app; 2 optional workshops; Measurements at baseline, 3-month, and 6-month visits	Counseled individually for 30-40 minutes on diet and exercise by a nurse practitioner during baseline visit. Healthy food plate, physical activity, and the importance of weight loss are key areas of focus during the counseling. Provided weighing scale	Weight change (kg), mean (SD), 3 months: C: –0.8 (2.1), I: –3.2 (3.1), between-groups *P*<.001, 6 months: C: –0.5 (2.9), I: –3.2 (4.1), between-groups *P*<.001. achieved 5% weight loss, 3 months: C: 8%, I: 25%, 6 months: C: 8%, I: 44%
Muralidharan et al [31] (2019), India, Indian, multicenter RCT	Received app at baseline visit. App offered reality television show video lessons to educate and encourage lifestyle behavior tracking and change. Video lessons highlighted challenges and suitable solutions. Automatic motivational messages offered according to user's progress, alerts to prompt tracking, quizzes to reinforce learning, and message function to chat with coaches. Coaches provided weekly calls to revisit topics and emailed reports	Baseline visit to download app, weekly coach calls, emails, and text messages. Remote coach support through app chat. Measurements at baseline and 12-week visit	Received standard care that included a brochure on healthy eating, weight loss, and exercise at baseline visit. Offered face-to-face counseling with nutritionist	Weight change (kg), mean, 3 months: I: –1.1, within-group *P*<.01, C: –0.3, within-group *P*=.05, between-groups: –0.8, *P*<.05, achieved 5% weight loss: I: 15%, C: 9%
Oh et al [32] (2015), South Korea, Korean, multicenter RCT	Received information on increasing physical activity and controlling diet habits. Provided mobile phones for remote monitoring, body composition monitors, and pedometers. Advised to weigh daily or minimally 3 times weekly. Data were transmitted to the central server for immediate feedback through the designed algorithm. Received weekly, monthly health reports on progress through app; Provided phone consultations by educated consultants on disease management, health education, exercise, medication, and proper nutrition	Baseline visit, 12-week, and 24-week measurement visits. Remote app support and phone consultations by educated consultants	Received basic information on increasing physical activity and controlling diet habits at baseline visit. Body weight scales and pedometers were provided along with body weight journal for self-recording of weight and waist circumference; Returned for 2 in-person visits for measurements, consultations with physicians, and received advice about their nutrition and exercise	Weight change (kg), mean (SD) 6 months: I: –2.21 (3.60), *P*<.001, C: –0.77 (2.77), *P*<.001, between-groups *P*<.001

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Shin et al [33] (2017), South Korea, Korean, pilot RCT	Received a 5-minute face-to-face education on diet and exercise from a nurse with standardized education material (1200 kcal sample menu), exercise recommendations, and behavior modification. Offered Fit.Life wireless physical activity tracker with Bluetooth transmission and detailed instructions on activity tracker use with demonstration and handouts. Provided clear activity goals and advised to track activity on app; Additional features for tracking progress to hit financial goals; Intervention I1: app, intervention I2: app+financial incentives	Baseline visit for 5-minute education, detailed demonstration, and instruction on app use; 4 measurement visits (baseline, week 4, week 8, and week 12)	Received a 5-minute face-to-face education on diet and exercise from a trained nurse. Content included the clinical consequence of obesity, a dietary recommendation for weight loss with an example of a 1200 kcal sample diet menu, and a physical activity recommendation with specification of frequency, intensity, time, and type	Weight change (kg), mean (SD) 3 months: I2 (app + financial incentives): –3.1 (3.7), I1 (app): –1.1 (2.9), C: –0.4 (2.5), *P* value between C and I2<.001, between C and I1=.38, and between I1 and I2=.006
Suen et al [34] (2019), Hong Kong, Chinese, feasibility RCT	Received coaching on applying auricular acupressure with instructions on frequency of application. Return demonstration was required to ensure proper treatment. Information booklet and mobile app provided. App provided daily reminders for self-pressing, encouraged compliance, and tracking of self-pressing and bowel movement; App provided relevant multimedia information and precaution on auricular acupressure. Users could communicate with researchers through app and were reminded of return visit dates on app	Researchers met patients twice weekly to change tapes. Remote communication and advice provided on app if patients had questions or problems	Received coaching on applying auricular acupressure with instructions on frequency of application. Return demonstration was required to ensure proper treatment; Information booklet provided. Patients were requested to manually record the frequency of daily pressing and bowel movement. Researchers met patients twice weekly to change tapes	Weight change (kg), mean, 8 weeks: C: –1.33, *P*=.005, I: –1.56, *P*<.001, no significant difference between groups
Tanaka et al [35] (2018), Japan, Japanese, RCT	Assigned to a group with up to 6 members where users could share meal photos in the group chat of the culturally tailored app, FiNC, and receive direct feedback, instructions, advice, and encouragement from a nutrition professional. Users could also communicate with other users for social support. Specific FiNC-method dietary recommendations were provided without any calorie restriction. Self-monitoring and group learning were encouraged on the app	Baseline, 8-week, and 12-week in-person measurements. No further human intervention. Remote communication with certified nutrition professional through app	Nil intervention provided during the 12-week waitlist period (Controls received intervention afterward for 8 weeks)	Weight change, mean (95% CI), 8 weeks: I: –1.4 (–2.0 to –0.8), C: –0.1 (–0.6 to 0.4), between-groups *P*=.001, week follow-up after active intervention, 12 weeks: I: –1.4 (–2.1 to –0.8), C: –0.1 (–0.7 to 0.6), between-groups *P*=.004

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Yang et al [49] (2017), Taiwan, Chinese, crossover RCT	Received mobile physical activity promotion tool inclusive of lifestyle counseling, professional personal counseling, constructive feedback, health information, individualized reminder message at least once a week through Line app and email. A self-monitoring app with mobile activity sensor was provided along with an interactive internet webpage where users could track their health, compare results with peers, and receive recommendations	Lifestyle counseling at baseline, minimal human contact; Remote coaching on app and sending of reminder messages at least once weekly through Line and email; Baseline, 12-week, and 24-week in-person measurement visits	Received lifestyle counseling and booklet containing health education in support of behavioral and educational advice for diet control, increased physical activity, less smoking and drinking, stress, and regular health examination. Information on the related risk factors, development and prevention of metabolic syndrome, and various websites were also provided to the patients	Weight change (kg), mean (SD), 6 months, with activity promotion system: pre–activity promotion system: 77.7 (15.1), post–activity promotion system: 76.4 (15.5), pre–usual care: 78.1 (16.6), post–usual care: 76.8 (15.8), between-groups *P*=.93; (Crossover study without washout period, nil results on specific phases of intervention)
Yang et al [36] (2020), South Korea, Korean, cluster RCT	Physicians provided education on the use of the medical instruments and smartphone app. Explained management targets and guidelines to patients. Provided glucometer, test strips, and electronic manometer monitoring. Users were asked to upload their daily SMBG[f] results, other biometric information, and weight through the app; Data automatically transmitted to the main server where physicians could check the results through a website and send feedback messages (praise, encouragement, feedback, and advice) at least once per week. Additional direct phone calls were conducted as required. Monthly face-to-face consultations offered	Baseline education on instruments, smartphone app, management targets and guidelines. Weekly feedback message through website or additional calls as necessary. Monthly in-person consultations to review progress, measurements, and management	Visited the private clinics and received face-to-face consultations every month for review and measurements	Weight change (kg), mean (95% CI), 3 months: I: −0.63 (−1.02 to −0.24), C: −0.88 (−2.65 to 0.90), adjusted mean difference to control: 0.22 (−1.26 to 1.71), between-groups *P*=.77
Zhang et al [37] (2019), China, Chinese, RCT	Group I1: app (basic), received basic diabetes education, including diet control, adequate exercise, SMBG, and regular review. Provided with glucose meter, test strips, and targets and encouraged to track BG[g], habits, and obtain diabetes-related knowledge through the app. Users could contact clinicians by phone or app; Group I2: app, interactive group I1 intervention+third-party professional diabetes health care team comprising a dietitian and a health manager. Health team provided feedback and recommendations on progress, BG, and lifestyle habits. Provided daily prompts (first month) and then monthly on the app by health care team and reviewed weekly glucose reports. Users were given BG targets and were able to contact clinicians by phone or app	Baseline, 3-month, and 6-month measurement visits, ability to contact clinicians on the web through app or phone. Group I1 received support from a clinician, group I2 had additional interactive support on app with web-based management health care team comprising a dietitian and a health manager	Provided basic education. Patients obtained diabetes-related knowledge and skills by self-learning and summarizing, and they adopted lifestyles and behaviors voluntarily. Equipped with a designated BG meter and test strips, patients were advised to record results in a logbook. They could contact clinicians through phone	Weight change (kg), mean (SD), baseline: C: 69.6 (10.0), I1 (app, basic): 72.3 (11.6), I2 (app with health team): 70.8 (11.9), 3 months: C: 69.6 (9.6), I1: 72.2 (11.9), I2: 70.9 (11.6), 6 months: C: 69.4 (9.9), I1: 72.0 (11.7), I2: 71.0 (11.6); There were no significant differences among the 3 groups for body weight at both 3 and 6 months

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Zhou et al [38] (2016), China, Chinese, pilot RCT	Downloaded the Welltang app at baseline visit and received diabetic knowledge on diet, exercise, medicine, blood glucose monitoring, and the latest guidelines for diabetes care. Users were asked to monitor 7-point finger blood glucose level for 1 day every 1-2 weeks (prompts in place) and track lifestyle habits. Advice on progress, values, target goals, and medication were offered by clinicians or study team through app once a week or fortnight. Users could communicate with clinicians through app, and an electronic action plan facilitated clinic review	Baseline, 1-month, 2-month, and 3-month in-person consultations. Remote interaction with clinicians on app as necessary. Weekly or fortnightly feedback from clinicians	Monthly visits to see physician to review blood glucose readings through logbooks. Patients were asked to monitor their 7-point finger capillary blood glucose level with a blood glucose meter 1-3 days before each clinic attendance to facilitate medication regimen adjustments	Weight change (kg), mean (SD), baseline: I: 62.4 (12.8), C: 62.5 (12.8), 3 months: I: 62.2 (11.0), C: 62.7 (12.1), *P* value not significant within or between groups
He et al [39] (2017), China, Chinese, cohort-based non-RCT	Received an official WeChat account for self-monitoring and immediate feedback on lifestyle habits. Users communicated and competed on weight loss progress. Users received scores for interactions, feedback information, or activity on the app, and top scorers were rewarded. Multimedia information on weight loss and an expert consulting group in place to answer questions	Nil baseline education or in-person session; 2 weight managers per work organization were trained to obtain participants' data on height, weight, and waist circumference before and after the interventions were initiated for both groups. Remote communication with experts through app	Routine publicity, such as the slogan "Take the stairs and lose weight," was provided to the control group. No further details specified	Weight change (kg), mean (SD): I: –2.09 (3.43), C: –1.78 (2.96), mean weight loss between the 2 groups for men was significant based on the stratification of age and educational level, weight loss changes were not significant for women
Kim et al [40] (2014), South Korea, Korean, matched controlled non-RCT	Baseline data recorded were transmitted to the app at first visit. Thereafter, users self-measured blood pressure and blood glucose levels, and data were automatically transmitted to hospital or medical staff through the app. Medical staff analyzed the data and sent tailored feedback to the patient once per week on average; App provided warning messages and advice when blood glucose levels were too high or too low. Study staff called users if they had hypoglycemia or no data were recorded	Nil baseline education. Measurement visits at baseline and 12 weeks. Remote supervision and weekly coaching on app. Staff called users if they had hypoglycemia or no data were recorded	Standard care, not clearly reported in paper	BMI change (kg/m^2), mean (SD), baseline: I: 25.0 (3.3), C: 24.9 (3.4), 3 months: I: 25.0 (3.4), within-group *P*=.80, C: 24.3 (3.1), within-group *P*=.06, no significant difference between groups

Author (year), country, ethnicity, study design	Intervention	Health staff involvement	Control treatment	Change in weight or weight-related outcomes
Kim et al [50] (2019), South Korea, Korean, cohort-based non-RCT	Received face-to-face counseling service at public health center from physician, nutritionist, exercise specialist, and nurse at baseline, 12 weeks, and 24 weeks. Offered activity monitors, sphygmomanometers, glucometers, body composition measuring devices, and app for self-monitoring. Instructed to sync activity at least 5 times weekly and upload meal pictures once a month. Remote weekly individualized service related to healthy lifestyles was provided by health professionals along with monthly reports. Access to web-based communities for each health center facilitated consultations; Received intensive nutritional consultations at health centers based on meal photos (Rewards such as mobile gift cards were offered to users with excellent performances, but this was not duly reported in the *Methods* section)	Baseline, 12-week, and 24-week consultations and measurements. Weekly individualized advice and services related to lifestyle habits provided by physicians, nurses, nutritionists, and physical activity experts who monitored health information on the web in real time; Intensive nutrition consultations at each visit to health center	After classification according to test results, tailored care plans were established. Face-to-face counseling services offered at public health center by team comprising a health manager (a health expert such as a physician or nurse), nutritionist, and certified exercise expert, who provided individual or group health consultation (consultations adhered to the 2011 One-Stop Health Service Consultation Manual)	BMI measured at baseline but not reported in results. Change in proportion of patients with metabolic risk factor (elevated waist circumference) according to Adult Treatment Panel III criteria: I: –62 male patients, I: –78 female patients, significant difference within group for both genders, C: –2 male patients, C: –20 female patients, significant difference within female group only. No significant difference between groups
Wijaya and Widiantoro [41] (2018), Taiwan, Indonesian, pretest-posttest design	The intervention group received Social Cognitive Theory skill-building by WGTC[h] for a 10-week program. The participants formed teams of 3 or 4 members based on friendship, received a booklet that provided physical activity–related knowledge, and were offered a 50-minute guidance on watch and app use at baseline. Individual and group performances were shown in the WGTC of the webpage where information was automatically transmitted by the iNCKU watches and the smartphone apps through daily use	Nil baseline education. Provision of items, booklet, guidance on use and purpose of study at baseline. Measurements at baseline and 10 weeks. No further human contact	Received explanation for the purposes of the study and a booklet that provided physical activity–related knowledge	Weight change (kg), mean (SD), baseline: I: 59.04 (65.44), C: 58.11 (66.83), 10 weeks: I: 57.78 (64.30), C: 57.92 (66.70), between-groups *P*=.64, intervention recorded lower body weight, *P*<.002

[a]RCT: randomized controlled trial.

[b]DPP: diabetes prevention program.

[c]I: intervention.

[d]C: control.

[e]SMART-CR/SP: Smartphone and Social Media–Based Cardiac Rehabilitation and Secondary Prevention.

[f]SMBG: self-monitoring blood glucose.

[g]BG: blood glucose.

[h]WGTC: web-based game with team competition.

The active intervention period ranged from 8 to 52 weeks, with a mean period of 18 weeks; the most common intervention periods were 6 months [25-28,30,32,37,39,49,50], followed by 3 months [29,31,33,35,36,38,40,47]. Of the 21 studies, only 1 study (5%) included a 4-week follow-up outcome measurement after the active intervention [35]. The average attrition rate was 11.5% (SD 8.53, range 0%-28%). Of the 21 studies, 2 (10%) consisted of single-component interventions using a mobile app exclusively [35,39], whereas 19 (90%) were multicomponent interventions incorporating additional components such as face-to-face consultations [25-30,32-34,36-38,47-50], reviews through phone calls or emails [26,27,29,31,32,36,40,48], supporting webpage [41], or financial incentives [33]. Of the

16 studies with face-to-face consultations, 4 (25%) also included tutorials to familiarize users on app use [29,30,36,41].

Across interventions, of the 21 studies, 12 (57%) used a researcher-designed app [27,28,30-34,36,40,41,49,50], whereas 9 (43%) used a commercially available app [25,26,29,35,37-39,47,48]. All studies had either used a culturally adapted app or had an app that was locally developed, except for 2 (10%) studies [29,47] that used a generic app developed in the United States. All studies reported using a multifunctional app, with the most common features being interactivity with health professionals; calorie, activity, or weight tracking; health and lifestyle information; and progress feedback (Multimedia Appendix 5 [25-41,47-50]). Other features such as gamification, which included scoring of app use to gain points for reward redemption and leader board positions, were less commonly reported. The average number of app features was 8.7 (SD 3.36, range 3-15). Comparing between commercially available and researcher-designed apps, the average numbers of app features were comparable at 7.7 (SD 3.15, range 3-12) and 9.5 (SD 3.45, range 5-15), respectively.

Risk-of-Bias Assessments

RCTs and Bias

The risk-of-bias assessments of the included studies are summarized in Tables 3 and 4 for RCTs and Table 5 for non-RCTs, respectively. More than half of the RCTs included in both the systematic review and meta-analysis were rated low risk for selection bias, attrition bias, reporting bias, and other biases. Of the 17 RCTs, 1 (6%) was rated high risk for random sequence generation because of errors in patient randomization [27] and 5 (29%) were rated unclear risk for allocation concealment because of insufficient details reported by the authors [27-29,37,38]. Performance bias was rated high risk for 53% (9/17) of the studies [26-30,32,33,35,38], whereas 29% (5/17) of the studies scored unclear risk [25,31,34,36,37]. Of the 17 studies, detection bias was rated unclear for most, except for 3 (18%) that were rated low risk [25,34,35]. Of the 17 studies, 3 (18%) were rated high risk for attrition bias, with attrition rates between 21% and 24% [27,28,31]; 2 (12%) were rated unclear risk [32,38]; and 1 (6%) was rated high risk for reporting bias because the authors did not report the primary outcome registered in the trial registration [31]. Apart from these bias domains, of the 17 studies, 4 (24%) and 3 (18%) were rated high risk and unclear risk with regard to other biases, respectively.

Table 3. Risk of bias within the randomized controlled trials for selection, performance, and detection bias domains (N=17).

Author, year, country	Selection bias (random sequence generation)	Selection bias (allocation concealment)	Performance bias	Detection bias (self-reported outcomes)	Detection bias (objective measures)
Bender et al [47], 2018, United States	Low risk	Unclear risk	High risk	Low risk	Low risk
Dong et al [48], 2018, China	Low risk	Unclear risk	High risk	Unclear risk	Unclear risk
Dorje et al [25], 2019, China	Low risk	Low risk	Unclear risk	Unclear risk	Low risk
Kaur et al [26], 2020, India	Low risk	Low risk	High risk	Unclear risk	Unclear risk
Kim et al [27], 2019, South Korea	High risk	Unclear risk	High risk	Unclear risk	Unclear risk
Lee et al [28], 2018, South Korea	Low risk	Unclear risk	High risk	Unclear risk	Unclear risk
Lee et al [29], 2019, South Korea	Low risk	Unclear risk	High risk	High risk	Unclear risk
Lim et al [30], 2020, Singapore	Low risk	Low risk	High risk	Low risk	Unclear risk
Muralidharan et al [31], 2019, India	Low risk	Low risk	Unclear risk	Low risk	Unclear risk
Oh et al [32], 2015, South Korea	Low risk	Low risk	High risk	Unclear risk	Unclear risk
Shin et al [33], 2017, South Korea	Low risk	Low risk	High risk	Unclear risk	Unclear risk
Suen et al [34], 2019, Hong Kong	Low risk	Low risk	Unclear risk	Unclear risk	Low risk
Tanaka et al [35], 2018, Japan	Low risk	Low risk	High risk	Unclear risk	Low risk
Yang et al [49], 2017, Taiwan	Low risk	Unclear risk	High risk	Unclear risk	Unclear risk
Yang et al [36], 2020, South Korea	Low risk	Low risk	Unclear risk	Unclear risk	Unclear risk
Zhang et al [37], 2019, China	Low risk	Unclear risk	Unclear risk	Low risk	Unclear risk
Zhou et al [38], 2016, China	Low risk	Unclear risk	High risk	Unclear risk	Unclear risk

Table 4. Risk of bias within the randomized controlled trials for attrition, reporting, other, and overall bias domains (N=17).

Author, year, country	Attrition bias	Reporting bias	Other bias	Overall bias
Bender et al [47], 2018, United States	High risk	Low risk	High risk	High risk
Dong et al [48], 2018, China	Low risk	Low risk	Low risk	Moderate risk
Dorje et al [25], 2019, China	Low risk	Low risk	Low risk	Low risk
Kaur et al [26], 2020, India	Low risk	High risk	Unclear risk	High risk
Kim et al [27], 2019, South Korea	High risk	Low risk	High risk	High risk
Lee et al [28], 2018, South Korea	High risk	Low risk	High risk	High risk
Lee et al [29], 2019, South Korea	Low risk	Low Risk	Low risk	High risk
Lim et al [30], 2020, Singapore	Low risk	Low risk	Low risk	Moderate risk
Muralidharan et al [31], 2019, India	High risk	Unclear risk	High risk	High risk
Oh et al [32], 2015, South Korea	Unclear risk	Low risk	Unclear risk	Moderate risk
Shin et al [33], 2017, South Korea	Low risk	Low risk	Unclear risk	Moderate risk
Suen et al [34], 2019, Hong Kong	Low risk	Unclear risk	High risk	Moderate risk
Tanaka et al [35], 2018, Japan	Low risk	Low risk	Low risk	Moderate risk
Yang et al [49], 2017, Taiwan	High risk	Unclear risk	High risk	High risk
Yang et al [36], 2020, South Korea	Low risk	Unclear risk	Low risk	Moderate risk
Zhang et al [37], 2019, China	Low risk	Unclear risk	Low risk	Moderate risk
Zhou et al [38], 2016, China	Unclear risk	Low risk	Low risk	Moderate risk

Table 5. Risk of bias within the non-RCTs[a] (N=4).

	Author, year, country			
	He et al [39], 2017, China	Kim et al [40], 2014, South Korea	Kim et al [50], 2019, South Korea	Wijaya and Widiantoro [41], 2018, Taiwan
Study design	Cohort-based non-RCT	Matched controlled non-RCT	Cohort based non-RCT	Pretest-posttest design
Bias due to confounding	Moderate risk	Serious risk	Critical risk	Moderate risk
Bias in selection of patients into the study	Low risk	Low risk	Low risk	Low risk
Bias in classification of interventions	Low risk	Low risk	Low risk	Low risk
Bias due to deviations from intended interventions	Moderate risk	Moderate risk	Serious risk	Low risk
Bias due to missing data	Low risk	Low risk	Low risk	Low risk
Bias in measurement of outcomes	Moderate risk	Serious risk	Moderate risk	Moderate risk
Bias in selection of the reported result	Moderate risk	Moderate risk	Moderate risk	Moderate risk
Overall bias	Moderate risk	Serious risk	Critical risk	Moderate risk

[a]RCT: randomized controlled trial.

Non-RCTs and Bias

The overall risk of bias for non-RCTs ranged from moderate to critical risk. All studies were rated low risk for selection bias, classification bias, and attrition bias. The studies were mostly rated moderate risk of bias for the other domains, except for 1 study that was rated serious risk for confounding bias and detection bias, leading to the overall rating of serious bias for the study [40]. A non-RCT that was excluded from the meta-analysis had an overall critical risk score because of critical risk of bias for confounding and serious risk of bias for deviation from the intended intervention [50].

Intervention Efficacy

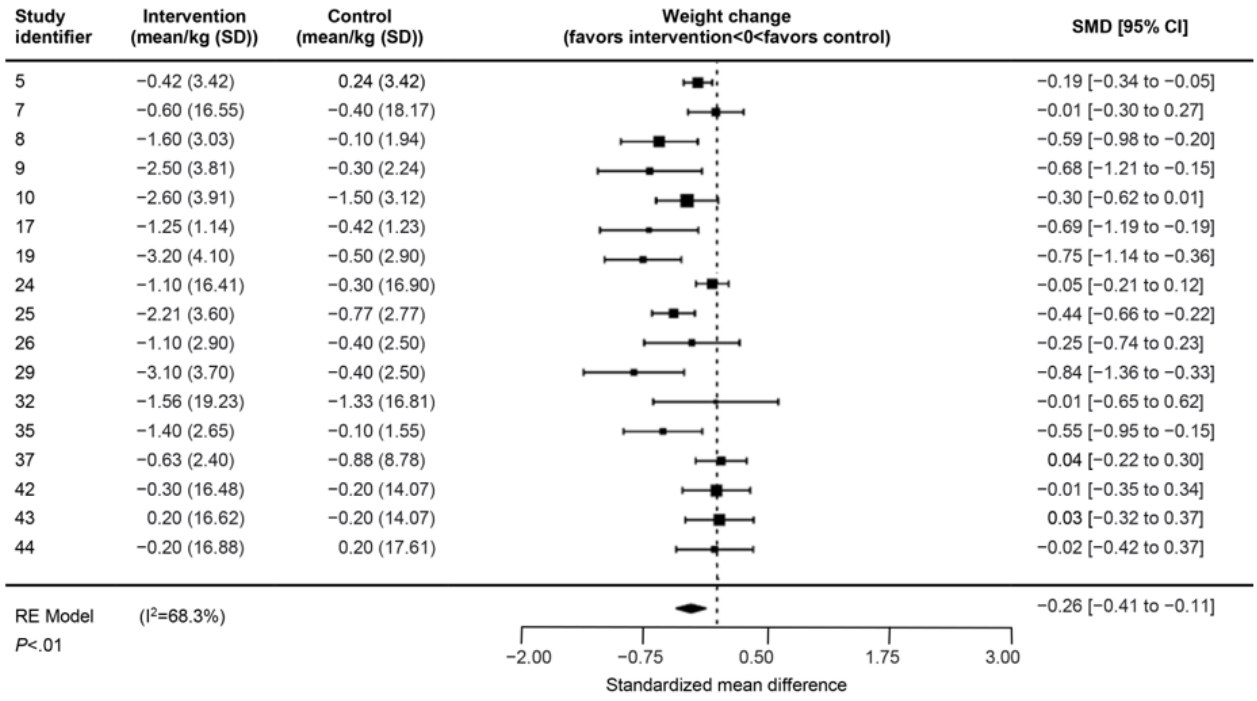
Of the 21 studies, 4 (19%; 3 RCTs and 1 non-RCT) were excluded from the meta-analysis because they did not provide specific values for weight-related outcomes that could be pooled [47-50]. In Bender et al [47], the authors reported significant

weight loss and higher percentage of intervention patients achieving 5% weight loss, whereas the remaining 3 studies did not report any significant results.

Among the 14 selected RCTs for the quantitative analysis, 9 (64%) reported a positive effect for the intervention on weight loss between the treatment arms [26,28-33,35,47]. In the 8 RCTs that did not report significant differences between the treatment groups, 7 (88%) nonetheless found that the intervention arms contributed to greater weight loss compared with the controls [27,34,38,39,41,49,50]. In all, 3 RCTs assessed the proportion of participants achieving clinically significant weight loss (ie, 5% weight loss) and found that between 15% and 44% of the participants in the intervention groups achieved 5% weight loss, whereas 6%-9% of the control participants achieved clinically significant weight loss [30,31,47].

The pooled weighted effect size across 14 RCTs for weight change (Figure 2) was small to moderate (Hedges g=–0.26; 95% CI –0.41 to –0.11; P<.01) with substantial heterogeneity (I^2=68.3%), whereas similar effect sizes were also observed for BMI (Hedges g=–0.21; 95% CI –0.42 to –0.01; P=.04; I^2=69.9%) and waist circumference (Hedges g=–0.24; 95% CI –0.45 to –0.02; P=.03; I^2=65.5%; Figures S1 and S2 of Multimedia Appendix 6). In terms of absolute weight, BMI, and waist circumference reduction, the raw mean difference revealed that the intervention group lost 1.16 kg (95% CI 0.81-1.52), 0.42 kg/m^2 (95% CI 0.16-0.68), and 1.21 cm (95% CI 0.22-2.21) more than the control group, respectively (Figures S3-S5 of Multimedia Appendix 6). Visual assessment of the funnel plots revealed no obvious asymmetry, suggesting a low risk of publication bias (Multimedia Appendix 7).

Figure 2. Forest plot showing the pooled effects of the interventions that incorporate apps on weight change. RE: random effects; SMD: standardized mean difference.

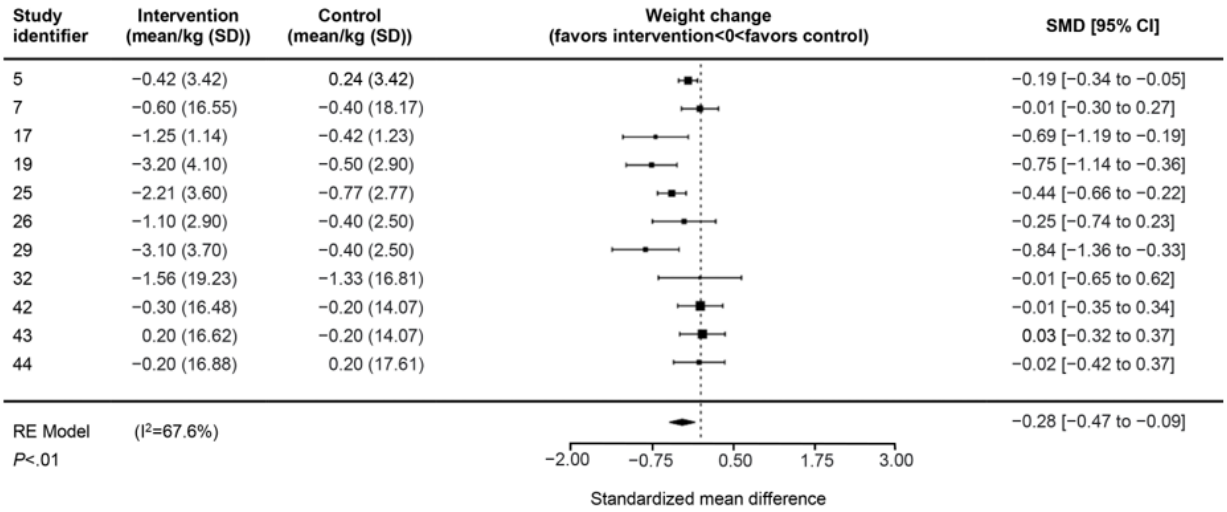


In a separate analysis of non-RCTs, 2 studies were included for each outcome: weight and BMI (Figures S6 and S7 of Multimedia Appendix 6). The effect size for weight change was statistically significant but small, with Hedges g=–0.09 (95% CI –0.13 to –0.05; P<.01; I^2=0%). In contrast, the effect size for BMI was not statistically significant (Hedges g=0.06; 95% CI –0.27 to 0.39; P=.74; I^2=0%). No analysis was conducted for waist circumference because only 1 study reported this outcome [39]. Overall, the results for the non-RCT meta-analysis should be interpreted with caution because there were very few data points included in the analyses and the data from the studies included were highly variable.

In the subgroup analyses for single-component (standalone app interventions) studies, no meta-analysis was conducted because of a lack of data points to assess the outcomes for RCTs and non-RCTs separately. Tanaka et al [35] reported a statistically significant weight loss between the groups but not He et al [39], although the intervention patients achieved a greater weight loss.

For interventions with the addition of apps to usual care, the effect size for weight (Figure 3) was statistically significant with a small to moderate Hedges g=–0.28 (95% CI –0.47 to –0.09; P<.01; I^2=67.6%); however, this was not the case for BMI and waist circumference outcomes (Figures S8 and S9 of Multimedia Appendix 6).

Figure 3. Forest plot showing the pooled effects of interventions combining usual care with app (intervention) versus usual care alone (control) on weight change. RE: random effects; SMD: standardized mean difference.



Moderation analysis suggested that study duration was not a significant moderator of intervention effects on weight, BMI, and waist circumference, with *P* values of .72, .67, and .69, respectively. All studies included in this meta-analysis had an intervention period of 6 months or less. Post hoc analyses (Figures S10 and S11 of Multimedia Appendix 6) revealed that changes in body weight were significantly different between the intervention and control groups, both in studies with a duration of 3 months or less (Hedges *g*=–0.28; 95% CI –0.52 to –0.05; *P*=.02; I^2=75%), as well as those between 3 and 6 months (Hedges *g*=–0.29; 95% CI –0.51 to –0.08; *P*=.01; I^2=72.2%). Conversely, changes in BMI and waist circumference did not differ significantly between the treatment arms in both data subsets (Figures S12-S15 of Multimedia Appendix 6).

Secondary Outcomes

Across all 21 studies, 17 (81%) reported at least one secondary outcome measured through a range of tools and approaches, as described in Multimedia Appendix 8 [25-41,47-50]. In all, 8 out of 11 studies (73%) showed that an app intervention improved dietary outcomes, with 5 studies [26,32,38,48,50] reporting concomitant within-group and between-groups differences, whereas Suen et al [34] and Lee et al [29] reported within-group differences. Physical activity outcomes were mostly reported as changes in exercise frequency, intensity, or duration. A total of 7 out of 11 studies (64%) showed an increase in physical activity level or exercise scores, with significant within-group and between-groups differences reported in 3 [29,38,50] and 6 studies [25,33,38,41,49,50], respectively. Significant improvements in self-efficacy scores were reported in 4 out of 6 studies (67%) [26,38,48,50], whereas 2 studies reported no significant change [27,41].

Among the 9 studies reporting significant weight change, 2 (22%) [26,32] reported significant improvements in diet, whereas 1 (11%) [33] reported significant increment in physical activity. Of the 9 studies reporting significant weight change, 3 (33%) studies did not find a significant improvement in diet [29,33,35], whereas the others did not report dietary outcomes [28,30,31,47]. Similarly, for physical activity, 3 (33%) studies [28,29,32] did not find any significant improvement, whereas the remaining studies did not report this outcome [26,30,31,35,47].

Of the 21 studies, 9 (43%) reported on app use and engagement [25,29,30,33,35,37-39,50]; however, only 6 offered use statistics, albeit by using various measures [25,30,35,37-39]. These statistics included number of daily log-ins [30]; meal, activity, or weight logging [30]; uploading of meal photos [35]; reading of messages [25]; frequency of app use [37]; interactivity with coaches [37,38]; and number of questions asked and cumulative app use scores [39]. Of the 9 studies, 4 (44%) reported that increased app use or adherence was associated with greater weight loss and health outcomes [29,33,35,39], whereas the remaining 5 (56%) did not explore any associations. In addition, of the 9 studies, 2 (22%) found that app use declined over time [30,35], whereas the remaining 7 (78%) did not report app engagement trends.

Discussion

Principal Findings

The interventions that incorporated apps produced a small to moderate effect in reducing weight, BMI, and waist circumference in Asian populations, although substantial heterogeneity was present. It was unclear if single-component standalone app studies were efficacious for weight loss; however, supplementing usual care with an app seemed to be beneficial for enhancing weight loss compared with usual care alone. However, the results may not be representative of long-term studies because of a lack of data. This review also found that app interventions may be beneficial for improving diet and increasing physical activity and self-efficacy for healthy behaviors. In these interventions, apps made for Asian populations were largely culturally adapted and multifunctional, with the most common app features being communication with health professionals and self-monitoring of behaviors and outcomes. Overall, the quality of the studies ranged from low to unclear risk of bias for most domains, apart from performance bias where most of the studies were graded high risk because

of the lack of blinding, which is challenging in interventions that incorporate apps. Therefore, the results should be interpreted with caution.

Among the studies in our review that reported a significant difference in either between groups or within group for weight loss favoring the intervention or a greater likelihood among the intervention groups for clinically significant weight loss, most were multicomponent. They typically included face-to-face consultations and reviews through phone calls or emails in addition to the app component.

Our review found that supplementing multicomponent usual care practices with an app was successful in achieving greater results. Reviews of studies in Western populations have observed similar findings [8,15]. This could be attributed to the provision of social support, accountability, and increased opportunities for patients to be reviewed beyond the confines of the app [51], thus underscoring the importance of additional components to raise contact frequency, enhance self-monitoring, and maximize outcomes.

Monitoring of weight, diet, and physical activity behaviors was a common feature of apps in the interventions reviewed. The addition of an app to usual care aligns with the understanding that self-monitoring can improve self-regulation of behaviors and weight loss [52,53]. By enhancing convenience for users, apps thus encourage more consistent self-monitoring [54] to promote treatment adherence and weight loss [15,42]. This review also observed that the outcomes of healthy eating and increased physical activity, which are key determinants of weight loss, tend to occur alongside improved self-efficacy for implementing healthy behaviors. As self-efficacy was previously found to be positively associated with self-monitoring, it is likely that increased self-monitoring may account for the behavioral improvements seen [55].

It was also apparent from our review that multifunctional, *all-in-one* apps were common among Asian interventions. The features included direct communication with health professionals through the app, in addition to functionalities for calorie, activity, and weight tracking as well as provision of health information and progress feedback, thereby matching the features found in effective digital weight loss interventions reported in a recent review [54].

In contrast with apps designed in Western countries, which tend to focus more on independent learning [8], apps designed for Asians frequently include accessibility to health professionals. It is plausible that having health coaches within apps reduces the barriers for Asians to seek health information, validation, and support from their clinicians [18,19,56], while conferring increased credibility to the coaching and support given to users [54,57]. App users recognize the benefit of health professionals such as dietitians providing support, particularly as they offer effective, evidence-based, culturally appropriate, and tailored dietary counseling to participants [58,59]. Correspondingly, most studies that employed a dietitian or nutrition-trained professional reported a greater weight loss with the intervention [28,30-32,35,37,50].

Furthermore, all but 2 studies [29,47] either used a culturally adapted app or an app that was locally developed; employed the respective country's native language; and incorporated localized educational content, food databases, and recommendations. Research supports that using culturally appropriate content, engaging local facilitators, and offering the app in the native language are important factors that may promote app use and outcomes [57,60].

Only 6 of the 21 studies reported app use statistics, whereas 2 others provided associations between app adherence and outcomes without reporting actual app engagement data. Meyerowitz-Katz et al [61] reported in a recent meta-analysis that the pooled estimate of app nonuse (defined as attrition rate) was 43%, indicating a serious limitation of app-based interventions if strategies for maintaining long-term engagement with the intervention (longer than a year) are not considered. In this review, the interventions that incorporated an app achieved statistical significance for weight change in studies with durations that were 6 months or less. However, the efficacy of these interventions in long-term studies remains unclear because none of the studies reviewed included durations longer than a year. App engagement levels in this review varied across studies, making comparison difficult. Nevertheless, evidence from this review echoed the results reported by previous reviews that increased app use is associated with greater adherence and weight loss [8,15], notwithstanding the fact that app engagement typically declines over time [8,61].

Strengths and Limitations

This is the first meta-analysis to report on the efficacy of apps incorporated into interventions targeting weight loss with or without healthy behavioral change in populations of Asian ethnicity. The review and meta-analysis were conducted according to best practice and followed PRISMA guidelines with a comprehensive search strategy and assessment of risk of bias using Cochrane Collaboration tools. The study selection, data extraction, and quality assessment were conducted independently by 2 reviewers.

This review is not without some limitations. The heterogeneity observed across studies was substantial, making it challenging to effectively interpret the results. Substantial heterogeneity could be due to the differences in study aims, targeted outcomes, methods, populations, and interventions. The lack of consistent and detailed reporting among the studies limited our ability to assess the true dose of intervention received, user engagement levels, and behavior change techniques that may have been employed in the apps. Therefore, the results of the meta-analysis should be interpreted with caution. Future studies that incorporate apps should consider using a standardized tool such as the Behavior Change Technique Taxonomy to code app features in a systematic and replicable manner and report user engagement statistics to evaluate app use and outcomes. This review was also limited to studies that were published in the English language. This may be problematic, given that we were studying apps in Asian populations, and English may not have been the first language in many countries; hence, some articles in other languages would not have been captured in this review. Finally, as most of the studies were multicomponent in nature,

components apart from the app, such as in-person education or review calls, may have more strongly influenced the outcomes; however, it was not possible to identify the contribution of these components to the weight loss outcomes.

Conclusions

This review contributes to the literature by presenting quantitative evidence that multicomponent interventions that incorporate apps produce a small to moderate effect toward weight loss in studies of Asian populations with intervention periods of 6 months or less. It is unclear if single-component standalone-app studies are efficacious for weight loss; however, adding apps to multicomponent usual care confers better outcomes. More evidence is required to determine the efficacy of apps in the long term. Cultural adaptation and offering the app in the native language of the users seem to be a priority in Asian apps. Multifunctional apps with features that include self-monitoring, individualized feedback, and a link to health professionals within the apps may be key to raising awareness and engagement, as well as promoting weight loss. Finally, it is recommended that investigators monitor and address the low uptake of apps and attempt to enhance engagement level before using apps as part of national health strategies for reducing obesity.

Acknowledgments

The authors wish to thank librarian Monica Cooper from the University of Sydney for her assistance with the development and refinement of the search strategy. The authors are also grateful for the support from the Singapore Ministry of Health's National Medical Research Council under its Health Services Research Grant program (NMRC/HSRG/0063/2016).

Authors' Contributions

SMA and JC jointly designed the research question and research protocol, and JHL, JJ, YYD, MAF, and SLL approved the protocol. SMA and JC performed the searches, screened the studies, extracted the data, and assessed quality. SMA and JJ conducted the gray literature searches, hand searching of reference lists, and extraction of data on app features from the studies. LJH conducted the meta-analysis and guided the quantitative data analysis. SMA and JC drafted the manuscript. All authors contributed to writing the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Modifications to PROSPERO protocol.
[PDF File (Adobe PDF File), 31 KB - jmir_v23i11e28185_app1.pdf]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 53 KB - jmir_v23i11e28185_app2.pdf]

Multimedia Appendix 3
Search strategy for MEDLINE.
[PDF File (Adobe PDF File), 121 KB - jmir_v23i11e28185_app3.pdf]

Multimedia Appendix 4
Unique identifiers for meta-analyses.
[PDF File (Adobe PDF File), 141 KB - jmir_v23i11e28185_app4.pdf]

Multimedia Appendix 5
App features.
[PDF File (Adobe PDF File), 178 KB - jmir_v23i11e28185_app5.pdf]

Multimedia Appendix 6
Additional meta-analysis and subgroup analyses.
[PDF File (Adobe PDF File), 921 KB - jmir_v23i11e28185_app6.pdf]

Multimedia Appendix 7
Funnel plots of meta-analysis.
[PDF File (Adobe PDF File), 176 KB - jmir_v23i11e28185_app7.pdf]

Multimedia Appendix 8
Secondary outcomes and app engagement data.
[PDF File (Adobe PDF File), 212 KB - jmir_v23i11e28185_app8.pdf]

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8]
2. Prevalence of overweight among adults, BMI ≥ 25, age-standardized estimates by country. World Health Organization. 2017. URL: https://apps.who.int/gho/data/node.main.A897A?lang=en [accessed 2020-09-21]
3. Wealthy but Unhealthy: overweight and obesity in Asia and the Pacific: trends, costs, and policies for better health. Tokyo, Japan: Asian Development Bank Institute; 2018.
4. Misra A, Jayawardena R, Anoop S. Obesity in South Asia: phenotype, morbidities, and mitigation. Curr Obes Rep 2019 Mar;8(1):43-52. [doi: 10.1007/s13679-019-0328-0] [Medline: 30729382]
5. Fan J, Kim S, Wong VW. New trends on obesity and NAFLD in Asia. J Hepatol 2017 Oct;67(4):862-873. [doi: 10.1016/j.jhep.2017.06.003] [Medline: 28642059]
6. World Health Organization. Global Action Plan For The Prevention And Control Of NCDs. Geneva, Switzerland: WHO Press; 2013.
7. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484. [doi: 10.1111/obr.12903] [Medline: 31353783]
8. Schoeppe S, Alley S, Van LW, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
9. Poushter J. Smartphone ownership rates skyrocket in many emerging economies, but digital divide remains. Pew Research Center. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-rates-skyrocket-in-many-emerging-economies-but-digital-divide-remains/ [accessed 2020-07-01]
10. The mobile economy Asia Pacific 2020. GSM Association. 2020. URL: https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/06/GSMA_MobileEconomy_2020_AsiaPacific.pdf [accessed 2020-07-01]
11. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
12. Chen J, Gemming L, Hanning R, Allman-Farinelli M. Smartphone apps and the nutrition care process: current perspectives and future considerations. Patient Educ Couns 2018 Dec;101(4):750-757. [doi: 10.1016/j.pec.2017.11.011] [Medline: 29169863]
13. mHealth app economics 2017 - current status and future trends in mobile health. Research2Guidance. 2017 Nov. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2020-07-01]
14. Khoury CF, Karavetian M, Halfens RJ, Crutzen R, Khoja L, Schols JM. The effects of dietary mobile apps on nutritional outcomes in adults with chronic diseases: a systematic review and meta-analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651. [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
15. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
16. Kim H, Seo K. Smartphone-based health program for improving physical activity and tackling obesity for young adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
17. Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
18. Zhao M, Esposito N, Wang K. Cultural beliefs and attitudes toward health and health care among Asian-born women in the United States. J Obstet Gynecol Neonatal Nurs 2010;39(4):370-385. [doi: 10.1111/j.1552-6909.2010.01151.x] [Medline: 20629925]
19. Ngo-Metzger Q, Massagli MP, Clarridge BR, Manocchia M, Davis RB, Iezzoni LI, et al. Linguistic and cultural barriers to care. J Gen Intern Med 2003 Jan;18(1):44-52 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20205.x] [Medline: 12534763]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]

21. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
22. Strain M. 1983 to today: a history of mobile apps. The Guardian. 2015. URL: https://www.theguardian.com/media-network/2015/feb/13/history-mobile-apps-future-interactive-timeline [accessed 2020-01-20]
23. 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. Br Med J 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
24. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
25. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):363-374. [doi: 10.1016/s2589-7500(19)30151-7]
26. Kaur J, Kaur M, Chakrapani V, Webster J, Santos JA, Kumar R. Effectiveness of information technology-enabled 'SMART Eating' health promotion intervention: a cluster randomized controlled trial. PLoS One 2020;15(1):e0225892 [FREE Full text] [doi: 10.1371/journal.pone.0225892] [Medline: 31923189]
27. Kim EK, Kwak SH, Jung HS, Koo BK, Moon MK, Lim S, et al. The effect of a smartphone-based, patient-centered diabetes care system in patients with type 2 diabetes: a randomized, controlled trial for 24 weeks. Diabetes Care 2019 Jan;42(1):3-9 [FREE Full text] [doi: 10.2337/dc17-2197] [Medline: 30377185]
28. Lee CH, Cheung B, Yi G, Oh B, Oh YH. Mobile health, physical activity, and obesity: subanalysis of a randomized controlled trial. Medicine (Baltimore) 2018 Sep;97(38):e12309 [FREE Full text] [doi: 10.1097/MD.0000000000012309] [Medline: 30235680]
29. Lee K, Kim H, Lee S, Ha H. Changes in weight and health-related behavior using smartphone applications in patients with colorectal polyps. J Nutr Educ Behav 2019 May;51(5):539-546. [doi: 10.1016/j.jneb.2019.02.002] [Medline: 30902428]
30. Lim SL, Johal J, Ong KW, Han CY, Chan YH, Lee YM, et al. Lifestyle intervention enabled by mobile technology on weight loss in patients with nonalcoholic fatty liver disease: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 13;8(4):e14802 [FREE Full text] [doi: 10.2196/14802] [Medline: 32281943]
31. Muralidharan S, Ranjani H, Anjana R, Jena S, Tandon N, Gupta Y, et al. Engagement and weight loss: results from the mobile health and diabetes trial. Diabetes Technol Ther 2019 Sep;21(9):507-513. [doi: 10.1089/dia.2019.0134] [Medline: 31184922]
32. Oh B, Cho B, Han MK, Choi H, Lee MN, Kang H, et al. The effectiveness of mobile phone-based care for weight control in metabolic syndrome patients: randomized controlled trial. JMIR Mhealth Uhealth 2015;3(3):e83 [FREE Full text] [doi: 10.2196/mhealth.4222] [Medline: 26293568]
33. Shin DW, Yun JM, Shin J, Kwon H, Min HY, Joh H, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
34. Suen L, Wang W, Cheng KK, Chua MC, Yeung JW, Koh WK, et al. Self-administered auricular acupressure integrated with a smartphone app for weight reduction: randomized feasibility trial. JMIR Mhealth Uhealth 2019 May 29;7(5):e14386 [FREE Full text] [doi: 10.2196/14386] [Medline: 31144666]
35. Tanaka K, Sasai H, Wakaba K, Murakami S, Ueda M, Yamagata F, et al. Professional dietary coaching within a group chat using a smartphone application for weight loss: a randomized controlled trial. J Multidiscip Healthc 2018;11:339-347 [FREE Full text] [doi: 10.2147/JMDH.S165422] [Medline: 30038502]
36. Yang Y, Lee EY, Kim H, Lee S, Yoon K, Cho J. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266 [FREE Full text] [doi: 10.2196/16266] [Medline: 32130172]
37. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app-based interactive management on glycemic control in chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019 Dec 09;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
38. Zhou W, Chen M, Yuan J, Sun Y. Welltang - a smart phone-based diabetes management application - improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]
39. 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 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
40. Kim H, Choi W, Baek EK, Kim YA, Yang SJ, Choi IY, et al. Efficacy of the smartphone-based glucose management application stratified by user satisfaction. Diabetes Metab J 2014 Jun;38(3):204-210 [FREE Full text] [doi: 10.4093/dmj.2014.38.3.204] [Medline: 25003074]

41. Wijaya YM, Widiantoro FX. Using team supports and gamified web-based approaches to attain recommended levels of physical activity among international Indonesian students. Indones Nurs J Edu Clin 2019 Apr 23;3(2):138. [doi: 10.24990/injec.v3i2.214]
42. Jacobs S, Radnitz C, Hildebrandt T. Adherence as a predictor of weight loss in a commonly used smartphone application. Obes Res Clin Pract 2017;11(2):206-214. [doi: 10.1016/j.orcp.2016.05.001] [Medline: 27292942]
43. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
44. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions, 2nd Edition. Hoboken, New Jersey, United States: Wiley-Blackwell; 2019.
45. Viechtbauer W. Conducting meta-analyses in with the package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
46. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://tinyurl.com/28d6uf53 [accessed 2020-09-17]
47. Bender MS, Cooper BA, Flowers E, Ma R, Arai S. Filipinos fit and trim - a feasible and efficacious dpp-based intervention trial. Contemp Clin Trials Commun 2018 Dec;12:76-84 [FREE Full text] [doi: 10.1016/j.conctc.2018.09.004] [Medline: 30294699]
48. Dong Y, Wang P, Dai Z, Liu K, Jin Y, Li A, et al. Increased self-care activities and glycemic control rate in relation to health education via Wechat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018 Dec;97(50):e13632 [FREE Full text] [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
49. Yang Y, Wang C, Wang J, Lin C, Yang YC, Wang J, et al. The effects of an activity promotion system on active living in overweight subjects with metabolic abnormalities. Obes Res Clin Pract 2017;11(6):718-727. [doi: 10.1016/j.orcp.2017.06.002] [Medline: 28729003]
50. Kim T, Lee Y, Yu E, Kim M, Yang S, Hur Y, et al. Effects of a mobile healthcare service provided by public health centers on practicing of health behaviors and health risk factors. Nutr Res Pract 2019 Dec;13(6):509-520 [FREE Full text] [doi: 10.4162/nrp.2019.13.6.509] [Medline: 31814926]
51. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
52. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
53. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
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 Dec;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
55. Kim H, Faw M, Michaelides A. Mobile but connected: harnessing the power of self-efficacy and group support for weight loss success through mhealth intervention. J Health Commun 2017 May;22(5):395-402. [doi: 10.1080/10810730.2017.1296510] [Medline: 28339324]
56. Kim H, Ray CD, Veluscek AM. Complementary support from facilitators and peers for promoting mHealth engagement and weight loss. J Health Commun 2017 Nov;22(11):905-912. [doi: 10.1080/10810730.2017.1373876] [Medline: 29125393]
57. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
58. Williams LT, Barnes K, Ball L, Ross LJ, Sladdin I, Mitchell LJ. How effective are dietitians in weight management? A systematic review and meta-analysis of randomized controlled trials. Healthcare (Basel) 2019 Feb 01;7(1):20 [FREE Full text] [doi: 10.3390/healthcare7010020] [Medline: 30717197]
59. Chen J, Berkman W, Bardouh M, Ng CY, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216 [FREE Full text] [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
60. Patel RM, Misra R, Raj S, Balasubramanyam A. Effectiveness of a group-based culturally tailored lifestyle intervention program on changes in risk factors for type 2 diabetes among Asian Indians in the United States. J Diabetes Res 2017;2017:2751980 [FREE Full text] [doi: 10.1155/2017/2751980] [Medline: 28168201]
61. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]

Abbreviations

RCT: randomized controlled trial

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by J Murphy, K Villinger; comments to author 10.05.21; revised version received 10.06.21; accepted 23.09.21; published 16.11.21.

Please cite as:
Ang SM, Chen J, Liew JH, Johal J, Dan YY, Allman-Farinelli M, Lim SL
Efficacy of Interventions That Incorporate Mobile Apps in Facilitating Weight Loss and Health Behavior Change in the Asian Population: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(11):e28185
URL: https://www.jmir.org/2021/11/e28185
doi:10.2196/28185
PMID:34783674

Review

Digital Interventions on Healthy Lifestyle Management: Systematic Review

Ayan Chatterjee[1], MEng; Andreas Prinz[1], PhD; Martin Gerdes[1], PhD; Santiago Martinez[2], PhD

[1]Department for Information and Communication Technologies, Centre for e-Health, University of Agder, Grimstad, Norway

[2]Department of Health and Nursing Science, Centre for e-Health, University of Agder, Grimstad, Norway

Corresponding Author:
Ayan Chatterjee, MEng
Department for Information and Communication Technologies
Centre for e-Health
University of Agder
Jon Lilletuns Vei 9
Grimstad, 4879
Norway
Phone: 47 94719372
Email: ayan.chatterjee@uia.no

Abstract

Background: Digital interventions have tremendous potential to improve well-being and health care conveyance by improving adequacy, proficiency, availability, and personalization. They have gained acknowledgment in interventions for the management of a healthy lifestyle. Therefore, we are reviewing existing conceptual frameworks, digital intervention approaches, and associated methods to identify the impact of digital intervention on adopting a healthier lifestyle.

Objective: This study aims to evaluate the impact of digital interventions on weight management in maintaining a healthy lifestyle (eg, regular physical activity, healthy habits, and proper dietary patterns).

Methods: We conducted a systematic literature review to search the scientific databases (Nature, SpringerLink, Elsevier, IEEE Xplore, and PubMed) that included digital interventions on healthy lifestyle, focusing on preventing obesity and being overweight as a prime objective. Peer-reviewed articles published between 2015 and 2020 were included. We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and a framework for an evidence-based systematic review. Furthermore, we improved the review process by adopting the Rayyan tool and the Scale for the Assessment of Narrative Review Articles.

Results: Our initial searches identified 780 potential studies through electronic and manual searches; however, 107 articles in the final stage were cited following the specified inclusion and exclusion criteria. The identified methods for a successful digital intervention to promote a healthy lifestyle are self-monitoring, self-motivation, goal setting, personalized feedback, participant engagement, psychological empowerment, persuasion, digital literacy, efficacy, and credibility. In this study, we identified existing conceptual frameworks for digital interventions, different approaches to provide digital interventions, associated methods, and execution challenges and their impact on the promotion of healthy lifestyle management.

Conclusions: This systematic literature review selected intervention principles (rules), theories, design features, ways to determine efficient interventions, and weaknesses in healthy lifestyle management from established digital intervention approaches. The results help us understand how digital interventions influence lifestyle management and overcome the existing shortcomings. It serves as a basis for further research with a focus on designing, developing, testing, and evaluating the generation of personalized lifestyle recommendations as a part of digital health interventions.

(J Med Internet Res 2021;23(11):e26931) doi:10.2196/26931

KEYWORDS

eHealth; digital intervention; lifestyle; obesity; challenges; mobile phone

Introduction

Overview

Deaths caused by lifestyle diseases are increasing rapidly compared with deaths caused by infectious disease [1-4]. Most lifestyle diseases arise from unhealthy and sedentary lifestyles, low nutritional propensities, and poor living conditions, affecting individuals from different financial backgrounds, beyond age and gender biases [1-3]. Lifestyle diseases are a monetary burden to individuals, families, businesses, and governments. They cause 41 million deaths each year, equivalent to 71% of all deaths worldwide [1-6]. Every year, 15 million people die from lifestyle diseases between the ages of 30 and 69 years, and more than 85% of these *premature* deaths occur in low- and middle-income countries [1-6]. The fundamental risk factors [7-13] behind lifestyle diseases are excessive alcohol and tobacco consumption, improper food plan, and physical inactivity, resulting in excess weight gain (obesity), increased blood glucose, hypertension, high blood cholesterol, and social detachment [7-13]. Obesity is a primary lifestyle disease that leads to other lifestyle diseases, such as cardiovascular diseases (CVDs), clinical obstructive pulmonary disease, cancer, type 2 diabetes, hypertension, and depression [7-13]. In 2016, more than 1.9 billion adults age ≥18 years were overweight. Of these, over 650 million were obese [2-4]. In 2016, 39% of adults aged ≥18 years were overweight, and 13% were obese [2-4]. In 2019, an estimated 38.2 million children aged ≤5 years were overweight or obese [2-4]. Once considered a problem in high-income countries, overweight and obesity are now rising in low- and middle-income countries (especially in urban environments) [2-4]. Since 2000, the number of overweight children aged ≤5 years has increased by nearly 24% in Africa [2-4,8]. In 2019, almost half of the children aged ≤5 years who were overweight or obese lived in Asia [2-4,8]. Thus, control plans need to include detection, screening, treatment, and prevention methods. Digital interventions may provide viable and hypothetically cost-effective models to improve well-being. They provide widely distributed, trusted, and personalized well-being information and services to fulfill individualized needs to maintain a healthy lifestyle [14-19].

Digital interventions for changing negative health behaviors to advance a healthy lifestyle are instigated by persuasion studies. The World Health Organization (WHO) has classified digital health interventions into the following 4 categories: clients, health care providers, health system managers, and data services, where digital and mobile technologies are being used to help well-being system needs and achieve health objectives [20-26]. Digital intervention methods include conceptualization, intervention strategies, policy design, understanding of the environment, motivation, behavioral determinants and psychology, persuasion, self-determination theory, self-regulation, participation (engagement), decision-making and feedback generation, goal setting and evaluation, incorporation of digital technologies (eg, smartphones, computers, and wearable sensors), and digital recommendation generation. Its success depends on credibility, satisfaction, privacy, digital literacy, proper connectivity, cocreation, and efficacy evaluation [20-26]. According to the WHO, harnessing the power of digital technology is critical to achieving universal health coverage, and digital technologies are not an end in themselves; they are essential tools for promoting health, maintaining world security, and serving the disadvantaged [18-24]. Digital interventions have been carried out effectively for well-being advancement and psychological well-being and for enabling self-administration of enduring conditions [18-24]. Nonetheless, their capability is restricted by low use rates, with noncommitment being a significant challenge. The use of digital technology provides new opportunities to improve health [18-24]. However, the evidence also highlights the challenges posed by certain interventions. If digital technologies are to be maintained and integrated into health systems, they must be able to demonstrate long-term improvements compared with traditional methods of providing health services [23,24].

Limitations

Digital intervention depends heavily on the context and ensures the proper design, including the structural issues in the environment in which they are used, the available infrastructure, the health needs they are addressing, and the ease of use of technology [1,15-17]. Consequently, it is important to discover successful procedures for expanding individual engagement with digital interventions. Digital interventions can also support health workers to give them more opportunities to clinical protocols around, for example, decision support mechanisms or telemedicine consultation [15-17]. Digital interventions should supplement and enhance the functions of the health system by accelerating information exchange and other mechanisms but cannot substitute the essential components required by the health system, such as the health workforce, funding, leadership and governance, and access to critical medicines. However, digital interventions in health care have tremendous potential as scalable tools to enhance well-being and health care service conveyance by improving viability, proficiency, openness, and personalization [1,15-17]. The WHO is working hard to ensure that it can be used as efficiently as possible, which means adding value to medical staff and individuals who use these technologies, considering the limitations of the infrastructure, and making appropriate coordination [22-24]. Although digital intervention's application area is broad, we have focused only on *digital behavioral intervention for healthy lifestyle* studies in this paper.

Study Aim

Efficiency, acceptability, and compliance are 3 necessary indicators of digital behavioral interventions, as they are prerequisites for positively affecting health or healthy behavior. Efficacy refers to the effect of using technology in behavioral interventions. Acceptability means that users are satisfied with the technology. Compliance refers to the degree to which technology is used, as expected. This systematic literature review addresses the following research questions (RQs):

RQ1: what are the existing conceptual frameworks for digital interventions for healthy lifestyle management?

RQ2: what are the different *approaches* to provide digital interventions for healthy lifestyle and what are the essential *methods*?

RQ3: what is the importance of digital intervention in promoting healthy lifestyles targeting obesity and overweight?

Methods

Overview

We used a systematic literature review method to obtain a broad overview of the current literature on the subject in a reproducible and understandable manner. A systematic review is a study of the evidence of a clearly expressed problem. It uses systematic and transparent strategies to understand, select, and strictly evaluate related basic research and extract and explore facts from the research covered in the evaluation. Systematic reviews represent scientific synthesis of evidence. We used the PRISMA [27] and Rayyan [28] evidence-based frameworks for systematic literature reviews. Subsequently, for article selection, we used the Scale for the Assessment of Narrative Review Articles (SANRA) scaling [29]. We conducted our systematic literature review following the Wendler [30] proposals to search scientific databases, as explained in *Strategy* subsection.

Strategy

The system's search strategy was designed using a combination of thesaurus, Oxford dictionary, and free terms covering the following terms: *eCoach, e-Coach, eHealth, e-Health, electroniccoaching, online, automatic, persuasion, persuasive technology, mHealth, mobilehealth, digital health, mobile, digitalintervention, smartphone, smart-phone, application, app, prevention, healthybehaviour, healthybehavior, behaviourchange, behaviorchange, exercise, activity, walk, step, fitness, sitting, inactive, screen time, sport, leisure activity, nutrition, nutritional, diet, dietary, healthy eating, salad, vegetables, fruit, discretionary food, snack, sweet beverage, carbonated beverage, soft drink, habit, tobacco, alcohol, computer, sedentary, lifestyle, intervention, lifestyle recommendation, digital recommendation, behavioral recommendation, program, programme, conceptual model, health promotion, prevention, obesity, overweight, weight-gain, weight gain, weight change, engagement, effectiveness, efficiency, credibility, trust, motivation, regulation, challenges, preferences, sample*, and *study duration*. We filtered our search with the following search string: ((*eCoach* OR **Coach* OR **coaching* OR *automatic* OR *online* OR *eHealth* OR *e-Health* OR *persuasion* OR *persuasive** OR *mHealth* OR *mobile** OR *digitalhealth*) AND (*digitalintervention* OR *smartphone* OR *smart-phone* OR *computer* OR *app** OR *prevention* OR **recommendation*) AND (**behavior** OR **behaviour** OR *sedentary** OR *lifestyle* OR *exercise* OR **activity* OR *walk* OR *step* OR *fitness* OR *inactive* OR *screentime* OR *sport* OR *nutrition** OR *diet** OR *healthyeating* OR *salad* OR *veg** OR *fruit* OR *discretionary** OR *snack* OR **beverage* OR *softdrink* OR *habit* OR *tobacco* OR *alcohol*) AND (*engagement* OR *persuasion* OR *effectiveness* OR *efficacy* OR *efficiency* OR **motivation* OR **regulation* OR *challenge** OR *limitation** OR *credibility* OR *trust* OR *preferences* OR *program* OR *programme* OR *conceptual** OR *sample* OR **duration*) AND (*obesity* OR **weight**)). We conducted the search in collaboration with the library of the University of Agder in Norway on the following 5 electronic databases including Nature, SpringerLink, Elsevier, IEEE Xplore, and PubMed, as they produced the maximum number of scientific sources related to digital intervention studies for healthy behavior targeting on the prevention of obesity and overweight as a primary objective. Related search keywords were identified using Medical Subject Headings terms, keywords from relevant articles, synonyms, and self-established search terms. *EndNote (V.9.x)*, *DOAJ*, *SHERPA/RoMEO*, and *Microsoft Excel (Office 365, 2019)* were used to search, collect, and select related articles effectively.

Studies on digital intervention are promising and have an enormous scope in the health care domain with information and communication technologies. A systematic literature review produced related older studies, but they are primarily in the theoretical phase, and the practical implementation is very young. Therefore, to keep our systematic literature review focused, articles related to digital interventions for healthy lifestyle management were included when published between January 1, 2015, and December 15, 2020. The search was limited to English literature, humans, digital health intervention methods, and research focused on improving healthy lifestyles. We aim to include peer-reviewed articles that describe digital intervention methodologies, conceptual models, theories, key challenges, lifestyle recommendations, and research related to healthy lifestyle management focusing on preventing obesity and being overweight with digital means. Articles are classified into the following groups: quantitative, qualitative, both quantitative and qualitative, and short papers, such as posters, editorials, and commentaries. Quantitative analysis is the factual examination of information gathered by the framework to test explicit speculations. Qualitative analysis centers around words and implications to investigate thoughts and encounters inside and out. The selection criteria (or specific parameters) for the quantitative and qualitative articles were (1) articles associated with a healthy lifestyle, with the main goal of preventing obesity and being overweight using digital interventions (or recommendations); (2) methods, theories, and strategies associated with digital interventions; and (3) challenges of digital interventions for lifestyle change.

We aim to adopt the explicit inclusion and exclusion criteria, as described in Textbox 1 and divide and distribute the articles among authors to complete the screening using the *Rayyan* collaboration and research tool. After individual screening, the results will be verified by other authors to resolve discrepancies between the reviewers. Subsequently, eligible peer-reviewed articles will be identified by manual search, quality score, and manual assessment of reference lists of related papers. Initially, titles, keywords, abstracts, and conclusions will be screened for inclusion. Then, we review the screened articles independently and check for individual eligibility for final inclusion.

Textbox 1. Inclusion and exclusion criteria for systematic literature review.

Inclusion criteria

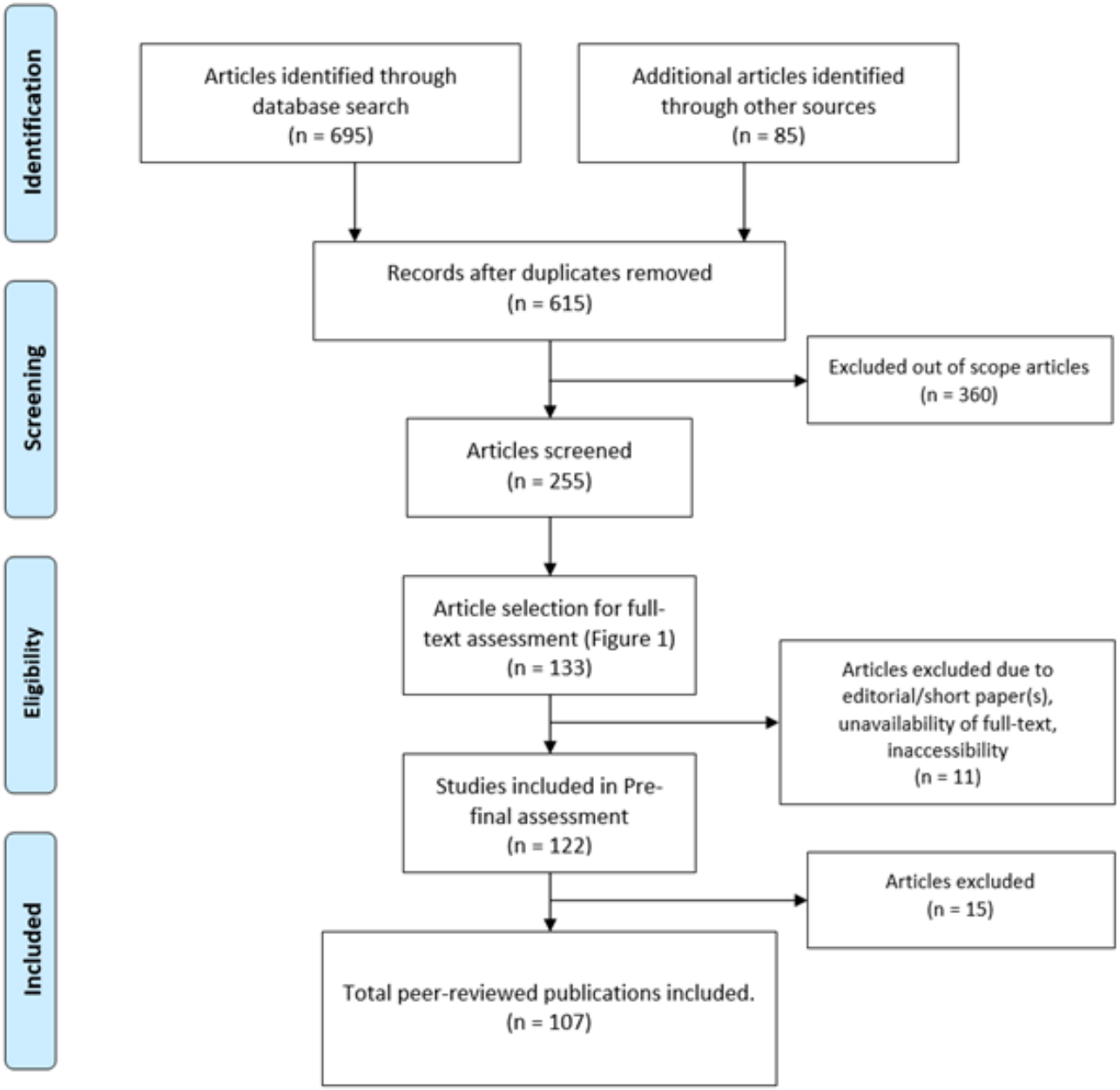
- Peer-reviewed, full-length articles written in English
- Digital intervention on healthy lifestyle articles published in the selected databases between 2015 and 2020
- Articles indexed in *Google Scholar* and (*SCOPUS* or *SCI* or *SCIE*)
- Journal papers, conference papers, or books
- Both qualitative (primary and secondary research) and quantitative studies
- *Open access* and accessible through the university library
- Studies associated with behavior change prevention rather than treatment and management of health conditions

Exclusion criteria

- Articles not written in English
- Incomplete, non–peer-reviewed articles
- Nonexperimental studies
- Poster, editorial, and commentary papers
- Articles published outside the selected time frame (<2015 and >2020)
- Articles not indexed in *Google Scholar* and (*SCOPUS* or *SCI* or *SCIE*)
- Studies related to offline human behavior interventions
- Papers having a considerable amount of analogous content or exact duplicate articles
- Economic, investment, and policy-making articles related to digital interventions
- Articles related to impacts from social interactions on behavior changes
- Articles related to traditional nutritional and physical activity assessment without any adoption of digital intervention methodologies
- Articles related to behavioral impacts on other lifestyle diseases apart from obesity such as cancer, mental health, chronic obstructive pulmonary disease, cardiovascular disease, dysglycemia, type 2 diabetes, loneliness, and hypertension
- Articles related to the treatment and management of health conditions rather than prevention of lifestyle disease (obesity and overweight in context)
- Articles related to cultural adaptation, nutrition policy, pregnancy, and genetics
- Articles related to worksite wellness, remote patient monitoring, hospital care, osteoporosis prevention for older adults, and web-based intervention for victims of cyberbullying

The search results from the databases for full-text assessment are as follows: IEEE Xplore (n=2), Nature (n=4), Elsevier (n=28), PubMed (n=40), and SpringerLink (n=59); “n” signifies total number articles screened for full-text assessment. We then excluded short research articles (1-2 pages) from the previous search list (Figure 1). In the prefinal assessment, for data extraction, we maintained an Excel spreadsheet with the following fields: title, reference in the American Medical Association format, author, population size, study duration, target group (children, adolescents, adults, and older adults), nature of the paper (review, conceptual or methodology, survey, and implementation), year, country of research, key terms, keywords, intervention type, publication channel, technology use, peer-reviewed, key findings (outcome, measures, intervention methods, theory, intervention components, effectiveness, and results), nature of assessment (qualitative, quantitative, or both), key challenges, and quality score based on SANRA. The primary outcome indicators extracted from the preliminary research results were digital intervention methods, nutritional intake, physical exercise, and healthy habits (consumption of tobacco and alcohol). The quality of the included articles was assessed using the SANRA 0-2-point scale. We graded individual papers based on the 6 quality parameters, as defined in Textbox 2. Individual quality parameters were subcategorized into a 0-2–point scale. Finally, we calculated the mean score of the 6 quality parameters. The number of studies included in the prefinal assessment was 122 with a SANRA score >1.90. Out of 122 studies, 15 had a SANRA score between 1.9 and 2.0. Therefore, we selected 107 articles that scored scale 2 based on the SANRA scale (Textbox 2).

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



Textbox 2. Searched article scaling based on quality parameters.

Quality parameters

- Justification of the article's importance for readership
- Statement of concrete or specific aims or formulation of questions
- Conference and journal papers to describe the overview of this study
- Referencing
- Scientific reasoning
- Appropriate presentation of data

In the final stage, we had articles with grade 2 and cited them as references (Textbox 3). In addition, we added 30 articles to the reference list (Textbox 3), including websites (accessed URLs), conference and journal papers to describe the overview of this study, risks of lifestyle diseases, and healthy behavior plans. The complete process of choosing the source for this study is shown in the flowchart in Figure 1. The article selection process consists of four stages—identification, screening, eligibility, and inclusion. Instead of depicting all the included studies as separate tables, we have presented significant findings from the respective studies.

Textbox 3. Nature of studies in the reference list.

Nature of studies

- Articles and web resources (URLs) to describe the overview of this study, risks of lifestyle diseases, and healthy behavior plan [2-31]
- Included studies with the Scale for the Assessment of Narrative Review Articles scale 2 (107 articles) [1,32-137]

Results

Study Selection and Study Characteristics

The searches (electronic database and manual databases) resulted in 780 papers (695 in electronic databases and 85 manually), of which 165 were duplicates. A total of 360 articles were excluded from the study following the exclusion criteria, and 255 articles were screened by reading abstracts, keywords, and conclusion sections. We selected 133 articles for full-text reading. We decided on 122 articles after a full-text review and checked the paper's length and the full-text availability in the prefinal stage. The final search produced 107 core peer-reviewed articles eligible for citation (4 from Nature, 49 from Springer, 23 from Elsevier, 1 from IEEE, and 30 from PubMed) related to digital intervention for healthy lifestyle management. Of the 107 papers, 72 (67.3%) were qualitative, 29 (27.1%) were quantitative, and the remaining 6 were both qualitative and quantitative. The final 107 papers were selected from 4 continents, including Asia, Europe, North America, and Oceania with the subsequent detailed search results as indicated by "n": Europe (n=56), North America (n=32), Oceania (n=16), Asia (n=2), Asia and North America (n=1), North America and Europe (n=1), and Oceania and North America (n=1).

The selected studies were clustered among the following 4 study groups that helped to answer our RQs: review (n=38), implementation (n=37), conceptual or methodology (n=23), and survey (n=9) studies related to digital interventions for healthy lifestyle management. Here, *n* signifies the total distribution of studies in the 4 study groups.

The findings related to the identified RQs are elaborated as follows.

RQ1: What Are the Existing Conceptual Frameworks for Digital Interventions for Healthy Lifestyle Management?

Mummah et al [32] proposed the concept of a conceptual framework with the following 10 phases: empathize with target users, define the target behavior, the basics of behavioral theory, come up with implementation strategies, potential prototype products, gather user feedback, build a real minimum product, a pilot test to evaluate potential efficacy and utility, evaluation of effectiveness in randomized controlled trials, and sharing of interventions and results. These phases are grouped into 4 overarching categories: integration, design, evaluation, and sharing.

Muench et al [33] proposed an overarching framework to perform digital triggering (text messages, emails, and push alerts) focusing on individual goals with the following 5 components: who (sender), how (stimulus type, delivery medium, and heterogeneity), when (delivered), how much (frequency and intensity), and what (trigger target, trigger structure, and trigger narrative). They showed how user characteristics, conceptual models, and clinical aims help to plan digital interventions and initiate tailoring with product features and user states.

Lewis et al [34] provided an idea to understand human behavior technology engagement to measure digital behavior change interventions (DBCIs) using a proposed framework. The proposed framework conceptualizes the 2 basic categories of commitment measured in digital behavior interventions (DBIs). The types are committed to health behaviors known as *Big E* and involvement of DBI known as *Small E*. DBI engagement has been further broken down into 2 subclasses: user interactions with intervention features designed to encourage frequent use, such as simple log-in, games, and social interactions and make the user experience attractive, and interactions of the user with the components of a behavior change intervention (ie, behavior change techniques) that influence determinants of health behaviors and then affect health behaviors.

Wang et al [35] proposed a holistic TUDER (Targeting, Understanding, Designing, Evaluating, and Refining) framework to integrate taxonomies into the theory-based digital health behavior intervention model. They showed how digital health behavior intervention is guided and influenced by theoretical concepts, such as behavior theories, behavior change technologies, and persuasive technology.

Lubans et al [36] proposed a framework for designing and delivering organized physical activity (PA) sessions for children and adolescents for effective dissemination. Recommended strategies include creating partnerships, presentations, intervention dissemination, scaling up research, and embedding evidence-based interventions.

According to Morgan et al [37], the limitations of digital health intervention programs include the lack of attention to critical sociocultural factors that affect participation and interventions on research results. Their research provides a conceptual model that illustrates the design and implementation of social and cultural interventions.

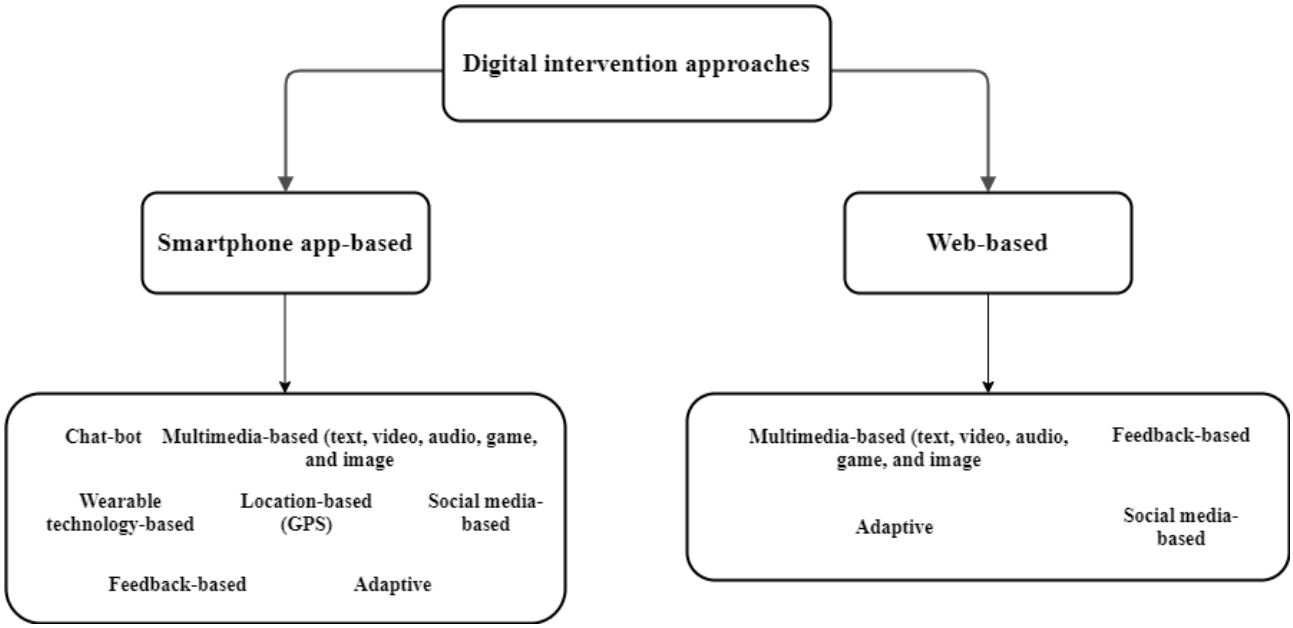
Hekler et al [38] proposed models and theories for DBCI based on international experts' discussions (including behavioral, computer, and health scientists and engineers) and provided suggestions for developing models and theories that can be learned from DBCI and can provide references. The proposed framework provides state-space representations to define when, where, for whom, and for the person in which the intervention will have a targeted effect. *State* refers to an individual's state based on various variables, which define the *space* in which an action mechanism may affect. The state-space representation can be used to help guide theorization and determine interdisciplinary methodological strategies to improve

measurement, experimental design, and analysis so that DBCI can match the complexity of real-world behavior changes.

RQ2: What Are the Different Approaches to Provide Digital Interventions for Healthy Lifestyle and What Are the Essential Methods?

In this systematic literature review, we identified the following 2 digital intervention approaches for a healthy lifestyle, and they are further structured in Figure 2: *smartphone app–based intervention (health monitoring and personalized recommendation generation)* [33,39-60,80-106] and *web-based intervention (web-based monitoring and self-management or self-reporting program)* [1,61-79,97-102].

Figure 2. Structuring of digital intervention approaches for a healthy lifestyle.



In this systematic literature review, 35 studies targeted app-based interventions, 18 targeted web-based interventions, and 19 targeted both app-based and web-based interventions. The essential methods associated with both types of intervention approaches are listed in Textbox 4.

Textbox 4. Essential methods associated with digital interventions for healthy lifestyle.

Study and key methods associated with digital interventions

- Lindwall et al [107]: motivation, satisfaction, self-determination, human-centered design, psychological needs (competence, autonomy, and relationship), and exercise-related behavior in selected motivational profiles
- Nicklas et al [108]: motivation and self-determination
- Yu et al [109]: self-monitoring for dietary intake—pen and paper documentation, self-check, and PDAs
- Karppinen et al [110]: motivation, persuasion with persuasive technologies, self-monitoring, goal setting, and evaluation
- Nouri et al [111]: user-centered design, observation, persuasive prompts, and feedback generation
- Sharpe et al [112]: efficacy evaluation, encouragement, and engagement
- O'Connor et al [113]: complexity analysis and identification of challenges
- Yardley et al [114]: feasibility study, usability study, beliefs, attitudes, needs, and conditions of people
- Fischer et al [105]: behavior change techniques, personal coaching, regular prompting, and efficacy evaluation
- Mitchell et al [53]: engagement, goal setting, incentives, and goal evaluation

RQ3: What Is the Importance of a Digital Intervention to Promote a Healthy Lifestyle Targeting Obesity and Overweight?

In recent years, an increasing number of digital intervention approaches have been implemented to promote a healthy lifestyle in different age groups. This RQ found modest evidence for effective digital interventions to improve PA, diet, and habits to prevent obesity and overweight. In individuals who are overweight and obese, therapeutic weight control approaches contribute to clinically significant weight losses; however, due to limited access, expense, and time constraints, many people cannot engage in these face-to-face treatments. The advancement of several digital weight loss services has resulted in

technological advances, such as universal access to the internet, expanded use of smartphones, and newer behavioral self-monitoring tools.

Verjans-Janssen et al [99] recognized the importance of implementing a long-term, locally relevant, holistic approach to promoting healthy weight status, stimulating the PA levels of children, and preventing them from wasting unnecessary time throughout school days on sedentary behaviors.

Brigden et al [115] designed interactive DBI for younger children based on the following characteristics: participation of parents, gaming functionality, additional therapist assistance, behavioral (rather than cognitive) approaches, and unique feedback and monitoring, shaping knowledge, repetition and substitution, and reward.

Nicklas et al [108] conceptualized a multi-exposure theory-based motivational theater, which can be an efficient behavior technique to improve preschool children's intake of vegetable dishes that can be conveniently disseminated to a large sample.

Lubans et al [36] used the Supportive, Active, Autonomous, Fair, and Enjoyable concepts to develop realistic strategies to engage young people with PA sessions to maximize involvement in PA and facilitate physical literacy by optimizing the results of affective, emotional, motivational, and movement skills.

Burrows et al [79] supported the need for web-based delivery of a balanced lifestyle program that addresses higher nutritional parental issues rather than infant weight. Parents were interested in a web-based family healthy lifestyle program and shared a desire for the program’s website to be easy to navigate and user-friendly, casual, but with personalized guidance and goal-setting opportunities.

Carrà et al [41] investigated *the Interactive Alcohol Risk Alertness Notifying Network for Adolescents and Young Adults (D-ARIANNA)*, a publicly accessible evidence-based eHealth app to estimate the current health risks by queries and fit-defined risk factors and include an overall risk score in percentage terms, accompanied by relevant images showing the main contributing factors in overview graphs and achievement.

Helle et al [57] conducted a study and identified 6 main behavioral risk factors as strong determinants of chronic diseases in adolescents (risky alcohol consumption, smoking, low diet, physical inactivity, sedentary behavior, and unhealthy sleep patterns). The study revealed that web and mobile technology interventions benefit adolescent participation, scope, and scalability to prevent the identification of health risk behaviors.

Stockwell et al [67] reported that PA and sedentary behavior are modifiable risk factors for lifestyle diseases and healthy aging; however, most of the older adults remain inadequately active. DBCIs can reach many older adults to promote PA and reduce sitting time. DBCIs may increase PA and physical function and reduce sedentary lifestyle and systolic blood pressure in older adults, but more high-quality testing is required.

According to Stephenson et al [40], machines, smartphones, and wearable technology resources can reduce the average sedentary time (minutes/day).

Weegen et al [62] showed that behavioral approaches were not successful without digital resources, and the integration of behavioral interventions with digital media proved to be an efficient way to stimulate PA.

Geidl et al [98] performed a recommendation generation study in adults with lifestyle diseases for PA and PA promotion over a week with a guideline of performing at least 150 minutes of aerobic PA with moderate intensity, 75 minutes of aerobic PA with vigorous intensity, or a combination of both. The PA and PA promotion guidelines advise adults impacted by lifestyle diseases and health providers on how much PA for adults with lifestyle diseases would be ideal. The guidelines provided the best strategies and approaches for growing low PA levels in adults with lifestyle diseases to professionals entrusted with PA promotion.

Gans et al [81] performed a study on 2525 worksite employees, and after 4 months, dietary fat intake decreased significantly with a multimedia-based (video) intervention strategy. Individually tailored videos helped office workers minimize dietary fat and increase fruit and vegetable consumption. Recently, to minimize sedentary behavior, technology-enhanced solutions such as mobile apps, activity monitors, prompting apps, SMS text messages, emails, and websites have been exploited.

Step-count sensors can improve walking, helping to tackle physical inactivity (pedometers, body-worn trackers, and smartphone apps). Chaudhry et al [87] assessed the influence of step-count monitors on PA in community-dwelling adults in randomized controlled trials, including longer-term results and discrepancies between step-count monitors and components of the intervention.

Muench et al [33] performed a multimedia-based (text) qualitative intervention to show the positive impact of digital triggers (such as SMS text messages, emails, and push alerts) in adults to change in curative conduct in health interventions. New technology apps for mobile health (mHealth) are emerging and provide the basis for fundamentally changing medical research, treatment practices, and scope.

Lin et al [43] conducted a web-based study on 4144 adults, collaborating with Quit Genius, an mHealth app focused on cognitive behavioral therapy that helps users quit smoking, to explore the successful nature of an mHealth digital app, which provides its users with substantial benefits and helps them modify their habits for a healthy lifestyle. The app's ability to improve users' hedonic well-being and inspire them mentally in their everyday lives was described as essential to help users quit smoking. The findings found that users whose well-being was improved via the app were 1.72 times more likely to quit smoking successfully.

Korinek et al [101] revealed that an adaptive phase target plus reward intervention using a mobile app appeared to be a feasible solution to increasing walking activity in overweight adults. Satisfaction with the app was strong, and the participants enjoyed having variable targets every day.

Mummah et al [32] tested the effect of a mobile app to increase vegetable consumption among overweight adults seeking to

sustain weight loss. The findings showed the effectiveness of a mobile app in increasing the consumption of vegetables among overweight adults.

Hanze University [104] launched a health promotion initiative to enable workers to lead a less sedentary life. The use of an activity tracker for tracking the regular step count of participants was one of the program's measures. For a fortnightly coaching session, the regular move count acted as feedback. They argued that the use of machine learning in the process of automated personalized coaching might become an invaluable advantage. Individualized algorithms allow PA to be predicted during the day and provide the ability to intervene in time. Machine learning techniques empower automatic coaching and personalization.

Since attending a weight control program, many individuals who are overweight find it difficult to sustain weight loss. Self-weighing and telephone support are useful tools for weight loss monitoring. Partridge et al [82,84] and Sidhu et al [83] tested the efficacy of a weight maintenance program based on SMS text messaging to facilitate daily self-weighing in adults and found it to be effective for young men and women.

Ball et al [69] organized an incentive-based, promising web-based intervention study to increase PA and reduce sitting among adults (ACHIEVE: Active Choices IncEntiVE). They explored the effectiveness, appeal, and impact of offering nonfinancial incentives for inactive middle-aged adults to encourage increased PA, decreased sedentary time, decreased BMI, and blood pressure.

Franssen et al [90] performed a study on consumer wearable activity trackers to promote PA levels.

Oosterveen et al [72] conducted a qualitative analysis of eHealth behavioral interventions aimed at analyzing smoking rates, nutritional habits, alcohol consumption, PA levels, and obesity in young adults and revealed that because of their high level of use of technology, eHealth interventions have potential among young adults.

Therefore, this RQ reveals that digital interventions have the potential to promote a healthy lifestyle (regular PA, healthy habits, and proper dietary intake) in all age groups, for personal weight management. In therapeutic approaches, self-monitoring is a crucial part of digital intervention [73]. According to a study [73], participants participating in behavioral interventions have lost 8%-10% of their initial body weight. These results are considered positive based on studies suggesting that losses of 5% can yield positive health improvements such as reductions in triglycerides, blood glucose, blood pressure, improved blood lipid levels, and a reduction in the risk of developing type 2 diabetes for a person. It is difficult to sustain long-term weight reductions achieved through behavioral therapy, and a different set of skills may be needed for success following interventions [1,32,42]. Many teenagers have low diet and PA patterns, which in later life can contribute to the development of lifestyle diseases. Web-based networks provide an affordable means of providing health interventions, but their efficacy is poorly understood. Investigation of the locations of PA and dietary patterns can promote *setting-specific* lifestyle interventions and increase knowledge of contextual vulnerabilities to poor health. As future directions for digital weight management, distribution, and policy implications should be emphasized [131-133].

Discussion

Research Questions

RQ1 helps to identify that behavioral theory, design thinking, evaluation, and the identification of limitations in the existing digital intervention frameworks are essential for successful, healthy lifestyle management. In contrast, *RQ2* reveals different approaches and methods associated with digital health interventions. Digital interventions for healthy lifestyle management have been categorized as discrete usefulness of advanced innovation applied to accomplish well-being goals and is executed inside digital well-being programs and information and communication technology frameworks, including communication channels, such as instant messages (SMS text message), alerts, and app-based notifications [116-118]. Digital intervention can motivate and stimulate individuals with self-tracking, goal setting, evaluation, and feedback or recommendation generation to promote a healthy lifestyle [118,119]. Different methods appear to impact health outcomes and usability. It would be interesting to test variants of component design and their impact on health outcomes and usability. Use of personalization to account for differences in preferences between groups of participants and even within groups of participants is essential in addition to cocreation between intervention developers and the target group [118-121]. Participants who attempted to self-manage their healthy lifestyle found that the most challenging part was to remain motivated [39]. They require apps that give them power and inspiration [39]. The study has confirmed that motivation is a multidimensional construct and people have different, sometimes competing, reasons for engaging in activities [39]. Moreover, human-centered analysis in digital intervention to study the intrinsic interactions of motivation and different regulations must be addressed. Despite the widespread use of mobile phones, digital literacy barriers are common among vulnerable people [39]. Participants have different participation levels in various activities, from higher to lower levels of participation. Researchers using traditional user-centered design methods should routinely measure these communication domains in their end-user samples. Future research should replicate these findings to a larger sample through direct observation, and persuasive prompts may be more effective in providing feedback to those with communication difficulties. *RQ3* summarizes the evidence on the importance of digital intervention that were exclusively directed at promoting healthy lifestyles, especially in children, adolescents, adults, and older adults. In the next section, we discuss the importance of digital intervention on healthy lifestyle promotion elaborately.

Digital Intervention and Healthy Lifestyle

A healthy lifestyle is a lifestyle that reduces the risk of severe illness or early death. Not all diseases are preventable, but a large proportion of deaths can be avoided, especially lifestyle or noncommunicable diseases. According to the *Harvard Medical School*, the key lifestyle factors to be monitored are

healthy diet, healthy PA level, healthy body weight, no tobacco consumption, and moderate alcohol intake [31]. According to *RQ3*, digital intervention can have a significant impact on healthy lifestyle management.

A study conducted by Steene et al [86] found significant country- and region-specific variations in PA and sedentary time in the European population, with lower PA levels. Boys in all age groups were more aggressive and less sedentary. At about 6 to 7 years of age, the initiation of age-related decline or leveling-off of PA and rise in sedentary time begins to become evident [86]. In children and adolescents, sedentary behavior strategies successfully decrease screen time; however, the scale of the effect tends to be limited [129]. The potential of digital intervention in older age groups outside of occupational settings and during sedentary leisure time must be examined in future studies. The sustainability of lifestyle changes in a positive direction remains a challenge [129]. Mobile apps to improve PA in young adults should include customized and personalized feedback and provide a coaching feature [58-60]. It is essential to create a well-oriented and easy-to-use interface with the ability to customize the app. The new area of mHealth is mHealth apps that target willing participants to enhance self-management of chronic conditions [58-60]. However, we found that only a small fraction of the mHealth apps available had been reviewed, and the amount of evidence was of inferior quality [58-60]. Improving the quality of evidence includes supporting prerelease app performance monitoring, designing few experiments, and performing better reviews with a rigorous risk of bias assessments [43,115]. Without enough evidence to back it up, for some time to come, digital intervention and app practicability will stall in their infancy [43,115,121]. Evidence suggests that an unhealthy lifestyle is associated with poor health outcomes [40,63]. It can have severe implications for health and well-being at any age, [63,64,126]. Therefore, there is a need to review the effects of multicomponent, complex interventions that include effective unhealthy lifestyle reduction strategies. We must focus on optimizing the effects of an intervention. Future intervention studies should use more rigorous methods to improve the quality of studies, considering larger sample sizes, randomized controlled designs, and valid and reliable lifestyle measurements. An overview of intervention development methods can help researchers understand various existing methods and comprehend the range of actions taken in intervention development before evaluating feasibility or pilot interventions [32,126,135].

One way to encourage PA and enhance health is to change the physical environment, but research on intervention efficacy is mixed [92-94]. Theoretical perspectives and conceptual problems are used in evaluative studies, and related literature can contribute to these inconsistencies [92-94]. Environmental and policy initiatives are socially incorporated into the framework and function through it. Therefore, a philosophical viewpoint must be considered and should be understood by evaluators. Future research should aim to explain how interventions function across disciplinary fields by considering these structures, the context in which interventions occur, and the measurable and unmeasurable mechanisms that might work [92-94,98]. It can be beneficial to promote health-based actions, such as PA, by using innovative and interactive media-based health education [92-94,98,134]. Therefore, to successfully influence behavior, it is essential to establish user-based techniques and reinforce the theories and hypotheses of behavioral change based on digital media. Step-count tracking [87] leads to improvements in short- and long-term step counts. There is no proof that either wearable sensors or smartphone apps, or extra counseling or incentives have additional benefits over more straightforward approaches focused on the pedometer [63,64]. To overcome the public health issue associated with physical inactivity, basic step-count tracking strategies should be prioritized. In general, it is not clear how self-reported sedentary behavior (eg, questionnaires, logs, and momentary ecological evaluations) compares with system measurement measures (eg, accelerometers and inclinometers) [63,64]. Evidence from this study indicates that when compared with system tests, single-item self-report measures typically underestimate sedentary time [63,64]. Therefore, to evaluate the reliability and validity of different self-report measures for evaluating sedentary activity, studies should exercise caution when comparing associations between various self-report and system measures with health performance.

In addition, video and adapting technologies have been effective in diet change measures; however, these methods have never been combined with researching personalized video efficacy [81]. Theory-based mobile interventions could provide a low-cost, scalable, and efficient approach to improving dietary habits and preventing associated chronic diseases. To encourage a healthy dietary pattern, nutrition messages or nutrient labeling, offering healthier choices, and portion size management of unhealthy foods have been potentially effective strategies in tertiary education environments [75]. The reduction in rates and the increased availability of nutritious choices in conjunction with nutrition knowledge have contributed to changes in dietary habits [75]. Further studies comparing the long-term efficacy of the climate and the combination of environmental policies to improve health outcomes are warranted. Dietary consumption has increased by increasing the availability of nutritious foods and reducing the portion size of unhealthy foods. In terms of modifying overall dietary patterns, the existing evidence base is misleading, as rising intake of desirable food groups was more effective than reducing unfavorable food habits, and fruit or vegetable intake and sugar-sweetened beverage consumption are the most notable observed changes [75,79]. Social support, followed by a demonstration of conduct, self-monitoring, goal setting, and feedback, is the most popular digital health behavior intervention [55,67,85,96]. In addition, a customized Facebook-based obesity prevention program for teenagers in Korea (Healthy Teens) [76] revealed usability problems in terms of material, appearance, and navigation. Facebook [76,96] has tremendous potential in promoting communication and engagement with immigrant teens, considering its prominence among adolescents. Interventions focused on social media (eg, Facebook) [66,96] are productive in facilitating meaningful improvements in adolescent eating habits. However, more research is required to explore effectiveness variations based on component tailoring, best use stimuli to promote behavior change over time, and keep people involved in changing

physical health behavior. The first step is to dismantle digital triggers into their parts and reassemble them according to their goals for improvement.

PA, sedentary time, and dietary habits vary across homes, schools, and other locations [95]. Health habits vary depending on the place or environment in which the participants are [95]. Although eating habits are typically more beneficial in home or school locations, PA is usually low and sedentary time in these locations is higher [95]. To optimize health habits in each area, digital interventions that address the various locations in which participants spend time and use location-specific behavior change techniques should be explored [95]. Among young people, binge drinking is prevalent [41]. eHealth technologies [41] are appealing to them and can be useful in raising awareness. However, to make eHealth apps suitable for longer-term effects, additional components, including daily feedback and repeated administration by different multimedia interventions, may be needed. Mass media campaigns [65] for smoking or tobacco programs are also effective over long periods. Digital interventions have been associated with decreased drinking and smoking frequency, with a slight yet persistent impact on teenagers and adults. Protective effects against alcohol and tobacco [65] use can be demonstrated through digital initiatives focused on a combination of social maturity and approaches to social influence. Evidence tends to be mixed with internet-based interventions, policy proposals, and incentives, and requires further study. Various distribution systems can enhance the effects of alcohol or tobacco misuse among teenagers and adults, including interactive platforms and policy initiatives. Adolescents are easily accessible by digital media and can represent a scalable and inexpensive opportunity to engage this audience in changing behavior [65]. Smartphone-based interventions [39-59] (such as apps, SMS text messages, sports, multicomponent interventions, emails, and social media) are readily available, inexpensive, and use tools already used by most teenagers. Therefore, it is essential to perform and publish high-quality academic literature studies and formally evaluate apps that have already been developed to inform the creation of potential interventions to change behavior. Essential improvements in behavior were also seen when interventions involved schooling, setting goals, self-monitoring, and parents' participation. Digital approaches [66] that include education, goal setting, self-monitoring, and parental participation can affect adolescents' meaningful health behavior changes. Most of the evidence relates to goal setting, further research into alternative media is needed, and it is essential to assess longer-term effects. There is a lack of evidence on the cost-effectiveness of digital health initiatives, and these data should be recorded in future trials. The young population has broadly embraced social media, so health researchers are searching for ways to exploit this social media involvement to deliver programs and health promotion campaigns [66,96]. In young adults, weight gain and suboptimal dietary choices are popular, and social media can be a possible instrument for encouraging and supporting healthy choices. The dissemination of information is now an appropriate use of social media by young adults. Careful evaluation is needed to use social media effectively for social support, either by private or by public sites, as its efficacy has yet to be demonstrated in experimental designs. In digital intervention studies aimed at manipulating weight, concerns about public social media use can lead to low engagement with social media [66,96,115]. Future research should explore how to use social media to better connect with young adults, how to use social media more efficiently to help young adults, and how to encourage social and peer-to-peer support to make healthy choices.

The systematic literature review has revealed that the identified digital intervention methods affect lifestyle behavior outcomes, focusing on PA, diet, habit, and associated primary and secondary health outcomes, such as fitness, motivation, reduced sedentary bouts, weight augmentation or weight status, blood pressure, glycemic responses, lipid profile, and quality of life in different study groups, as explained in the next section. Effectiveness (n=55), conceptualization (n=24), selection of appropriate methodology (n=25), effective and technological engagement (n=24), selection of strategies or techniques (n=20), motivation (n=15), feedback generation (n=15), environment (n=10), satisfaction (n=6), credibility (n=6), digital literacy (n=6), self-determination (n=5), and user-centric design (n=4) were responsible for the successful implementation of digital interventions for healthy lifestyle management. Here, *n* signifies the total number of overlapped studies in which the respective parameters are identified. Digital interventions [84,106,124] focused on mobile phone apps may be an acceptable and efficient way of encouraging weight loss in people who are overweight or obese. Digital health coaching can be a revolutionary approach to reduce barriers to access to much-needed weight loss therapies for obesity, given the ubiquity of mobile phones.

Strength and Limitation

A systematic literature review revealed that a healthy diet, healthy habits, and regular PA are powerful tools for reducing obesity and associated health risks. These findings bolster the use of digital interventions as a preventive option for obesity and overweight. Therefore, behavior change should be given the highest preference to avoid severe health damage. Planned digital interventions may potentially change growing negative behavior in humans with the adoption of persuasion, observation, goal evaluation, evidence-based personalized recommendation generation, health risk predictions (decision-making), automation, motivation, pragmatism, and trust. Developing and maintaining an empathetic relationship is perhaps the most critical determinant of successful digital intervention [106,124]. It is essential to know the participant first, and the interaction aspects are challenging owing to the delay in reaction time (both ways). Health care professionals need to ensure both relationship communication and goal-oriented coaching when using such digital intervention solutions. In the future, the quality of the interaction between the system and the participant will require attention if participants are to fully benefit from collaboration in digital intervention programs.

Digital intervention for healthy lifestyle management has great potential as a scalable tool that can improve health and health care delivery by improving effectiveness, efficiency, accessibility, security, and personalization. Therefore, a

knowledge base must be accumulated to provide information for developing and deploying digital health interventions [116,118]. However, the evaluation of digital health interventions poses unique challenges. Methodological limitations, selection of appropriate intervention methods, evaluation of efficacy, limitations of research on different populations, loss to follow-up, attrition rate, lack of participation in tracking, financial incentives and intervention burdens (long term or short term), digital literacy, technical participation, personalization, useful evidence-based automatic tailored lifestyle recommendation generation (intervention design), research heterogeneity, meta-analysis, cost-effectiveness (technical and financial feasibility analysis), trial selection, trial recruitment, scalability, accessibility, ethics, policy development, cyberbullying, safety, trust, user-centeredness, adoption of health care and collaboration methods that promote cooperation, unsustainable growth in complexity, and efficacy evaluation are some of the existing limitations of digital health interventions that should be overcome in existing research [126,128]. Although new technologies and rapidly changing technologies pose many unsolved problems, the broad consensus is that successful intervention design requires user-centered iterative development methods, hybrid methods, and in-depth qualitative research to gradually improve interventions to satisfy users. Therefore, conceptual participation (effective engagement) is essential to understand the relationship between the involvement in digital interventions and required behavioral changes and to achieve population-level benefits. Interventions must be delivered effectively at scale. Small effect sizes and high dropout rates [78,126] often affect web-based computer-tailored interventions, particularly among people with a low education level. The results and attractiveness of these remedies can theoretically be enhanced by using videos as a delivery format. The most successful and most appreciated intervention was the web-based video version of the computer-tailored obesity prevention intervention. Future research needs to analyze whether the results are sustained in the long run and how to maximize the intervention.

Conclusions

Digital intervention in health care is the intersection of health care, behavior science, computing, and engineering research and requires methods borrowed from all these disciplines. Digital interventions have effectively improved many health conditions and health behaviors; besides, they are increasingly being used in different health care fields, including self-management of long-term conditions, prevention of lifestyle diseases, and health promotion. In low-resource primary care environments, digital health strategies can be useful for preventing obesity. To minimize obesity and chronic disease risk among medically vulnerable adults in the primary care environment, digital health intervention uses an advanced digital health approach. The lack of user involvement hinders the full potential of digital interventions. There is an urgent need to develop effective strategies to promote user participation in digital interventions. One potential method is to use technology-based reminders or personalized recommendation generation. Compared with no strategy, technology-based strategies can promote participation. However, the findings of this systematic literature review should be understood with prudence, as only a few qualified studies have been identified for review, and the results are heterogeneous. The number and dates of studies indicate that a digital health intervention strategy is an emerging field. More research is needed to understand what strategic features are useful, their cost-effectiveness, and their applicability to different age groups. The results of this literature review will help to understand the concepts and parameters behind different DBI methods, thereby developing, testing, and evaluating the performance of a useful digital intervention in the future.

Acknowledgments

The authors (AC and AP) thank the coauthors (MG and SM) for reviewing the paper and providing useful comments to improve its quality. This research is unique, original, and has not been published or submitted elsewhere.

Authors' Contributions

The authors (AC and AP) divided and distributed the articles to complete the screening using the Rayyan tool. After individual screening, the results were verified by other authors (MG and SM) to resolve discrepancies between the reviewers.

Conflicts of Interest

None declared.

References

1. Foley P, Steinberg D, Levine E, Askew S, Batch BC, Puleo EM, et al. Track: A randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials 2016 Dec;48:12-20 [FREE Full text] [doi: 10.1016/j.cct.2016.03.006] [Medline: 26995281]
2. Chatterjee A, Prinz A, Gerdes M, Martinez S. An automatic ontology-based approach to support logical representation of observable and measurable data for healthy lifestyle management: Proof-of-concept study. J Med Internet Res 2021 Apr 09;23(4):e24656 [FREE Full text] [doi: 10.2196/24656] [Medline: 33835031]

3. Afshin A, Murray C, Reitsma M. Health effects of overweight and obesity in 195 countries over 25 years. Institute for Health Metrics and Evaluation. 2017. URL: http://www.healthdata.org/research-article/health-effects-overweight-and-obesity-195-countries-over-25-years [accessed 2021-05-01]
4. Toebes B, Hesselman M, van Dijk JP, Herman J. Curbing the lifestyle disease pandemic: Making progress on an interdisciplinary research agenda for law and policy interventions. BMC Int Health Hum Rights 2017 Sep 18;17(1):25 [FREE Full text] [doi: 10.1186/s12914-017-0131-5] [Medline: 28923078]
5. Traversy G, Chaput J. Alcohol Consumption and Obesity: An Update. Curr Obes Rep 2015 Mar 8;4(1):122-130 [FREE Full text] [doi: 10.1007/s13679-014-0129-4] [Medline: 25741455]
6. Bush T, Lovejoy JC, Deprey M, Carpenter KM. The effect of tobacco cessation on weight gain, obesity, and diabetes risk. Obesity (Silver Spring) 2016 Sep 29;24(9):1834-1841 [FREE Full text] [doi: 10.1002/oby.21582] [Medline: 27569117]
7. Vancampfort D, Koyanagi A, Ward PB, Rosenbaum S, Schuch FB, Mugisha J, et al. Chronic physical conditions, multimorbidity and physical activity across 46 low- and middle-income countries. Int J Behav Nutr Phys Act 2017 Jan 18;14(1):6 [FREE Full text] [doi: 10.1186/s12966-017-0463-5] [Medline: 28100238]
8. Obesity and overweight. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-05-01]
9. Klepac Pogrmilovic B, Varela AR, Pratt M, Milton K, Bauman A, Biddle SJ, et al. National physical activity and sedentary behaviour policies in 76 countries: Availability, comprehensiveness, implementation, and effectiveness. Int J Behav Nutr Phys Act 2020 Sep 18;17(1):116 [FREE Full text] [doi: 10.1186/s12966-020-01022-6] [Medline: 32948193]
10. Ashdown-Franks G, Vancampfort D, Firth J, Smith L, Sabiston CM, Stubbs B, et al. Association of leisure-time sedentary behavior with fast food and carbonated soft drink consumption among 133,555 adolescents aged 12-15 years in 44 low- and middle-income countries. Int J Behav Nutr Phys Act 2019 Apr 23;16(1):35 [FREE Full text] [doi: 10.1186/s12966-019-0796-3] [Medline: 31014359]
11. Sylvester B, Jackson B, Beauchamp M. The effects of variety and novelty on physical activity and healthy nutritional behaviors. Adv Motivat Sci 2018;5:169-202. [doi: 10.1016/bs.adms.2017.11.001]
12. Stubbs B, Vancampfort D, Firth J, Schuch FB, Hallgren M, Smith L, et al. Relationship between sedentary behavior and depression: A mediation analysis of influential factors across the lifespan among 42,469 people in low- and middle-income countries. J Affect Disord 2018 Mar 15;229:231-238. [doi: 10.1016/j.jad.2017.12.104] [Medline: 29329054]
13. Chatterjee A, Gerdes MW, Martinez SG. Identification of Risk Factors Associated with Obesity and Overweight-A Machine Learning Overview. Sensors (Basel) 2020 May 11;20(9):- [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
14. Labrique A, Agarwal S, Tamrat T, Mehl G. WHO Digital Health Guidelines: A milestone for global health. NPJ Digit Med 2020 Sep 18;3(1):120 [FREE Full text] [doi: 10.1038/s41746-020-00330-2] [Medline: 34535745]
15. Kolasa K, Kozinski G. How to Value Digital Health Interventions? A Systematic Literature Review. Int J Environ Res Public Health 2020 Mar 23;17(6):2119 [FREE Full text] [doi: 10.3390/ijerph17062119] [Medline: 32209988]
16. Chaput J, Willumsen J, Bull F, Chou R, Ekelund U, Firth J, et al. 2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5-17 years: Summary of the evidence. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):141 [FREE Full text] [doi: 10.1186/s12966-020-01037-z] [Medline: 33239009]
17. Tronieri JS, Wadden TA, Chao AM, Tsai AG. Primary care interventions for obesity: Review of the evidence. Curr Obes Rep 2019 Jun 19;8(2):128-136 [FREE Full text] [doi: 10.1007/s13679-019-00341-5] [Medline: 30888632]
18. Dempsey PC, Biddle SJ, Buman MP, Chastin S, Ekelund U, Friedenreich CM, et al. New global guidelines on sedentary behaviour and health for adults: Broadening the behavioural targets. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):151 [FREE Full text] [doi: 10.1186/s12966-020-01044-0] [Medline: 33239026]
19. González-Morales R, Canto-Osorio F, Stern D, Sánchez-Romero LM, Torres-Ibarra L, Hernández-López R, et al. Soft drink intake is associated with weight gain, regardless of physical activity levels: The health workers cohort study. Int J Behav Nutr Phys Act 2020 May 12;17(1):60 [FREE Full text] [doi: 10.1186/s12966-020-00963-2] [Medline: 32398158]
20. Weman-Josefsson K, Lindwall M, Ivarsson A. Need satisfaction, motivational regulations and exercise: Moderation and mediation effects. Int J Behav Nutr Phys Act 2015 May 20;12(1):67 [FREE Full text] [doi: 10.1186/s12966-015-0226-0] [Medline: 25990492]
21. Puggina A, Aleksovska K, Carlin A, Condello G, Cortis C, Jaeschke L, et al. "Determinants of Diet and Physical Activity" (DEDIPAC): An umbrella systematic literature review: Anna Puggina. Eur J Public Health 2016;26(1):ckw165.022. [doi: 10.1093/eurpub/ckw165.022]
22. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]
23. Classification of digital health interventions v1.0: A shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2021-05-01]
24. WHO releases first guideline on digital health interventions. World Health Organization. 2019. URL: https://www.who.int/news/item/17-04-2019-who-releases-first-guideline-on-digital-health-interventions [accessed 2021-05-01]
25. Chatterjee A, Gerdes MW, Martinez S. eHealth initiatives for the promotion of healthy lifestyle and allied implementation difficulties. In: Proceedings of the International Conference on Wireless and Mobile Computing, Networking and

Communications (WiMob). 2019 Presented at: International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); Oct. 21-23, 2019; Barcelona, Spain. [doi: 10.1109/wimob.2019.8923324]
26. Chatterjee A, Gerdes M, Prinz A, Martinez S. Human coaching methodologies for automatic electronic coaching (eCoaching) as behavioral interventions with information and communication technology: systematic review. J Med Internet Res 2021 Mar 24;23(3):e23533 [FREE Full text] [doi: 10.2196/23533] [Medline: 33759793]
27. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
28. 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]
29. Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. Res Integr Peer Rev 2019;4:5 [FREE Full text] [doi: 10.1186/s41073-019-0064-8] [Medline: 30962953]
30. Wendler R. The maturity of maturity model research: A systematic mapping study. Info Softw Technol 2012 Dec;54(12):1317-1339. [doi: 10.1016/j.infsof.2012.07.007]
31. Tello M. Healthy lifestyle: 5 keys to a longer life. Harvard Health Publishing. 2020. URL: https://www.health.harvard.edu/blog/healthy-lifestyle-5-keys-to-a-longer-life-2018070514186 [accessed 2021-05-01]
32. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
33. Muench F, Baumel A. More than a text message: Dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]
34. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: Pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
35. Wang Y, Fadhil A, Lange J, Reiterer H. Integrating taxonomies into theory-based digital health interventions for behavior change: A holistic framework. JMIR Res Protoc 2019 Jan 15;8(1):e8055 [FREE Full text] [doi: 10.2196/resprot.8055] [Medline: 30664477]
36. Lubans DR, Lonsdale C, Cohen K, Eather N, Beauchamp MR, Morgan PJ, et al. Framework for the design and delivery of organized physical activity sessions for children and adolescents: Rationale and description of the 'SAAFE' teaching principles. Int J Behav Nutr Phys Act 2017 Feb 23;14(1):24 [FREE Full text] [doi: 10.1186/s12966-017-0479-x] [Medline: 28231794]
37. Morgan PJ, Young MD, Smith JJ, Lubans DR. Targeted health behavior interventions promoting physical activity: A conceptual model. Exerc Sport Sci Rev 2016 Apr;44(2):71-80. [doi: 10.1249/JES.0000000000000075] [Medline: 26829248]
38. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
39. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: A taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
40. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: A systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
41. Carrà G, Crocamo C, Bartoli F, Carretta D, Schivalocchi A, Bebbington PE, et al. Impact of a mobile e-health intervention on binge drinking in young people: The digital-alcohol risk alertness notifying network for adolescents and young adults project. J Adolesc Health 2016 May;58(5):520-526. [doi: 10.1016/j.jadohealth.2016.01.008] [Medline: 27107907]
42. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
43. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: Exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
44. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
45. Kozak AT, Buscemi J, Hawkins MA, Wang ML, Breland JY, Ross KM, et al. Technology-based interventions for weight management: Current randomized controlled trial evidence and future directions. J Behav Med 2017 Feb;40(1):99-111. [doi: 10.1007/s10865-016-9805-z] [Medline: 27783259]
46. Mummah S, Robinson TN, Mathur M, Farzinkhou S, Sutton S, Gardner CD. Effect of a mobile app intervention on vegetable consumption in overweight adults: A randomized controlled trial. Int J Behav Nutr Phys Act 2017 Dec 15;14(1):125 [FREE Full text] [doi: 10.1186/s12966-017-0563-2] [Medline: 28915825]

47. Peyman N, Rezai-Rad M, Tehrani H, Gholian-Aval M, Vahedian-Shahroodi M, Heidarian Miri H. Digital Media-based Health Intervention on the promotion of Women's physical activity: A quasi-experimental study. BMC Public Health 2018 Jan 15;18(1):134 [FREE Full text] [doi: 10.1186/s12889-018-5025-5] [Medline: 29334970]
48. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 01;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
49. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
50. Järvelä-Reijonen E, Karhunen L, Sairanen E, Muotka J, Lindroos S, Laitinen J, et al. The effects of acceptance and commitment therapy on eating behavior and diet delivered through face-to-face contact and a mobile app: A randomized controlled trial. Int J Behav Nutr Phys Act 2018 Feb 27;15(1):22 [FREE Full text] [doi: 10.1186/s12966-018-0654-8] [Medline: 29482636]
51. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
52. Ahmed M, Oh A, Vanderlee L, Franco-Arellano B, Schermel A, Lou W, et al. A randomized controlled trial examining consumers' perceptions and opinions on using different versions of a FoodFlip© smartphone application for delivery of nutrition information. Int J Behav Nutr Phys Act 2020 Feb 12;17(1):22 [FREE Full text] [doi: 10.1186/s12966-020-0923-1] [Medline: 32050996]
53. Mitchell M, Lau E, White L, Faulkner G. Commercial app use linked with sustained physical activity in two Canadian provinces: A 12-month quasi-experimental study. Int J Behav Nutr Phys Act 2020 Feb 25;17(1):24 [FREE Full text] [doi: 10.1186/s12966-020-00926-7] [Medline: 32098625]
54. Pearson E, Prapavessis H, Higgins C, Petrella R, White L, Mitchell M. Adding team-based financial incentives to the Carrot Rewards physical activity app increases daily step count on a population scale: A 24-week matched case control study. Int J Behav Nutr Phys Act 2020 Nov 19;17(1):139 [FREE Full text] [doi: 10.1186/s12966-020-01043-1] [Medline: 33208166]
55. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
56. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: A systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
57. Helle C, Hillesund ER, Omholt ML, Øverby NC. Early food for future health: A randomized controlled trial evaluating the effect of an eHealth intervention aiming to promote healthy food habits from early childhood. BMC Public Health 2017 Sep 20;17(1):729 [FREE Full text] [doi: 10.1186/s12889-017-4731-8] [Medline: 28931384]
58. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: A systematic review. Pediatr Obes 2015 Dec;10(6):403-409. [doi: 10.1111/ijpo.12002] [Medline: 25641770]
59. Simons LP, Foerster F, Bruck PA, Motiwalla L, Jonker CM. Microlearning mApp raises health competence: Hybrid service design. Health Technol (Berl) 2015 Feb;5(1):35-43. [doi: 10.1007/s12553-015-0095-1] [Medline: 26097799]
60. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
61. Smith N, Liu S. A systematic review of the dose-response relationship between usage and outcomes of online physical activity weight-loss interventions. Internet Interv 2020 Dec;22:100344 [FREE Full text] [doi: 10.1016/j.invent.2020.100344] [Medline: 32995302]
62. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: A cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
63. Prince SA, Melvin A, Roberts KC, Butler GP, Thompson W. Sedentary behaviour surveillance in Canada: Trends, challenges and lessons learned. Int J Behav Nutr Phys Act 2020 Mar 10;17(1):34 [FREE Full text] [doi: 10.1186/s12966-020-00925-8] [Medline: 32151285]
64. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: A systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Mar 04;17(1):31 [FREE Full text] [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
65. Das JK, Salam RA, Arshad A, Finkelstein Y, Bhutta ZA. Interventions for adolescent substance abuse: An overview of systematic reviews. J Adolesc Health 2016 Oct;59(4S):61-75 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.06.021] [Medline: 27664597]

66. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
67. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: A systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
68. Garnett C, Crane D, Brown J, Kaner E, Beyer F, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: Meta-regression. J Med Internet Res 2018 Feb 28;20(2):e69 [FREE Full text] [doi: 10.2196/jmir.8807] [Medline: 29490895]
69. Ball K, Hunter RF, Maple J, Moodie M, Salmon J, Ong K, et al. Can an incentive-based intervention increase physical activity and reduce sitting among adults? the ACHIEVE (Active Choices IncEntiVE) feasibility study. Int J Behav Nutr Phys Act 2017 Dec 21;14(1):35 [FREE Full text] [doi: 10.1186/s12966-017-0490-2] [Medline: 28320409]
70. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Teeriniemi A, Salonurmi T, et al. Opportunities and challenges of behavior change support systems for enhancing habit formation: A qualitative study. J Biomed Inform 2018 Aug;84:82-92 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.012] [Medline: 29936091]
71. Teeriniemi A, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018 Nov 04;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]
72. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
73. 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]
74. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of ehealth and mhealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
75. Roy R, Kelly B, Rangan A, Allman-Farinelli M. Food environment interventions to improve the dietary behavior of young adults in tertiary education settings: A systematic literature review. J Acad Nutr Diet 2015 Oct;115(10):1647-1681. [doi: 10.1016/j.jand.2015.06.380] [Medline: 26271691]
76. Park BK, Nahm E, Rogers VE, Choi M, Friedmann E, Wilson M, et al. A facebook-based obesity prevention program for Korean American adolescents: Usability evaluation. J Pediatr Health Care 2017;31(1):57-66. [doi: 10.1016/j.pedhc.2016.02.002] [Medline: 26952300]
77. Byrne JL, Cameron WT, Maximova K, Browne NE, Holt NL, Cave AJ, et al. A brief eHealth tool delivered in primary care to help parents prevent childhood obesity: A randomized controlled trial. Pediatr Obes 2016 Nov 10:- [FREE Full text] [doi: 10.1111/ijpo.12200] [Medline: 27863165]
78. Walthouwer MJ, Oenema A, Lechner L, de Vries H. Comparing a video and text version of a web-based computer-tailored intervention for obesity prevention: A randomized controlled trial. J Med Internet Res 2015;17(10):e236 [FREE Full text] [doi: 10.2196/jmir.4083] [Medline: 26481772]
79. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition interventions for prevention and management of childhood obesity: What do parents want from an ehealth program? Nutrients 2015 Dec 15;7(12):10469-10479 [FREE Full text] [doi: 10.3390/nu7125546] [Medline: 26694456]
80. Williamson C, Baker G, Mutrie N, Niven A, Kelly P. Get the message? A scoping review of physical activity messaging. Int J Behav Nutr Phys Act 2020 Apr 15;17(1):51 [FREE Full text] [doi: 10.1186/s12966-020-00954-3] [Medline: 32295613]
81. Gans KM, Risica PM, Dulin-Keita A, Mello J, Dawood M, Strolla LO, et al. Innovative video tailoring for dietary change: Final results of the Good for you! Cluster randomized trial. Int J Behav Nutr Phys Act 2015 Oct 07;12(1):130 [FREE Full text] [doi: 10.1186/s12966-015-0282-5] [Medline: 26445486]
82. 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(1):7 [FREE Full text] [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]
83. Sidhu MS, Daley A, Jolly K. Evaluation of a text supported weight maintenance programme 'Lighten Up Plus' following a weight reduction programme: Randomised controlled trial. Int J Behav Nutr Phys Act 2016 Feb 12;13:19 [FREE Full text] [doi: 10.1186/s12966-016-0346-1] [Medline: 26867588]
84. Partridge SR, McGeechan K, Bauman A, Phongsavan P, Allman-Farinelli M. Improved eating behaviours mediate weight gain prevention of young adults: Moderation and mediation results of a randomised controlled trial of TXT2BFiT, mHealth program. Int J Behav Nutr Phys Act 2016 Apr 02;13:44 [FREE Full text] [doi: 10.1186/s12966-016-0368-8] [Medline: 27039178]

85. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MS. Social media use for nutrition outcomes in young adults: A mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
86. Steene-Johannessen J, Hansen BH, Dalene KE, Kolle E, Northstone K, Møller NC, Determinants of DietPhysical Activity knowledge hub (DEDIPAC); International Children's Accelerometry Database (ICAD) Collaborators, IDEFICS ConsortiumHELENA Consortium. Variations in accelerometry measured physical activity and sedentary time across Europe - harmonized analyses of 47,497 children and adolescents. Int J Behav Nutr Phys Act 2020 Mar 18;17(1):38 [FREE Full text] [doi: 10.1186/s12966-020-00930-x] [Medline: 32183834]
87. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: Systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
88. Dohrn I, Welmer A, Hagströmer M. Accelerometry-assessed physical activity and sedentary time and associations with chronic disease and hospital visits - a prospective cohort study with 15 years follow-up. Int J Behav Nutr Phys Act 2019 Dec 09;16(1):125 [FREE Full text] [doi: 10.1186/s12966-019-0878-2] [Medline: 31818303]
89. Chaix B, Benmarhnia T, Kestens Y, Brondeel R, Perchoux C, Gerber P, et al. Combining sensor tracking with a GPS-based mobility survey to better measure physical activity in trips: Public transport generates walking. Int J Behav Nutr Phys Act 2019 Oct 07;16(1):84 [FREE Full text] [doi: 10.1186/s12966-019-0841-2] [Medline: 31590666]
90. Franssen WM, Franssen GH, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
91. van Ekris E, Wijndaele K, Altenburg TM, Atkin AJ, Twisk J, Andersen LB, International Children's Accelerometry Database (ICAD) Collaborators. Tracking of total sedentary time and sedentary patterns in youth: A pooled analysis using the International Children's Accelerometry Database (ICAD). Int J Behav Nutr Phys Act 2020 May 18;17(1):65 [FREE Full text] [doi: 10.1186/s12966-020-00960-5] [Medline: 32423404]
92. Remmers T, Van Kann D, Kremers S, Ettema D, de Vries SI, Vos S, et al. Investigating longitudinal context-specific physical activity patterns in transition from primary to secondary school using accelerometers, GPS, and GIS. Int J Behav Nutr Phys Act 2020 May 18;17(1):66 [FREE Full text] [doi: 10.1186/s12966-020-00962-3] [Medline: 32423411]
93. Gupta N, Dencker-Larsen S, Lund Rasmussen C, McGregor D, Rasmussen CD, Thorsen SV, et al. The physical activity paradox revisited: A prospective study on compositional accelerometer data and long-term sickness absence. Int J Behav Nutr Phys Act 2020 Jul 20;17(1):93 [FREE Full text] [doi: 10.1186/s12966-020-00988-7] [Medline: 32690043]
94. Petersen TL, Brønd JC, Kristensen PL, Aadland E, Grøntved A, Jepsen R. Resemblance in accelerometer-assessed physical activity in families with children: The Lolland-Falster Health Study. Int J Behav Nutr Phys Act 2020 Dec 04;17(1):161 [FREE Full text] [doi: 10.1186/s12966-020-01067-7] [Medline: 33276796]
95. Ortega A, Bejarano CM, Cushing CC, Staggs VS, Papa AE, Steel C, et al. Differences in adolescent activity and dietary behaviors across home, school, and other locations warrant location-specific intervention approaches. Int J Behav Nutr Phys Act 2020 Sep 29;17(1):123 [FREE Full text] [doi: 10.1186/s12966-020-01027-1] [Medline: 32993715]
96. 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 Nov;63(5):531-545. [doi: 10.1016/j.jadohealth.2018.06.009] [Medline: 30197198]
97. Hermsen S, Frost J, Renes R, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. Comput Hum Behav 2016 Apr;57(5):61-74. [doi: 10.1016/j.chb.2015.12.023]
98. Geidl W, Abu-Omar K, Weege M, Messing S, Pfeifer K. German recommendations for physical activity and physical activity promotion in adults with noncommunicable diseases. Int J Behav Nutr Phys Act 2020 Feb 05;17(1):12 [FREE Full text] [doi: 10.1186/s12966-020-0919-x] [Medline: 32024526]
99. Verjans-Janssen SR, Gerards SM, Kremers SP, Vos SB, Jansen MW, Van Kann DH. Effects of the KEIGAAF intervention on the BMI z-score and energy balance-related behaviors of primary school-aged children. Int J Behav Nutr Phys Act 2020 Aug 17;17(1):105 [FREE Full text] [doi: 10.1186/s12966-020-01012-8] [Medline: 32807194]
100. 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 2016 Sep 23:446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
101. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: A longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
102. 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]

103. 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]
104. Dijkhuis TB, Blaauw FJ, van Ittersum MW, Velthuijsen H, Aiello M. Personalized physical activity coaching: A machine learning approach. Sensors (Basel) 2018 Feb 19;18(2):623 [FREE Full text] [doi: 10.3390/s18020623] [Medline: 29463052]
105. Fischer X, Donath L, Zwygart K, Gerber M, Faude O, Zahner L. Coaching and prompting for remote physical activity promotion: Study protocol of a three-arm randomized controlled trial (Movingcall). Int J Environ Res Public Health 2019 Jan 25;16(3):331 [FREE Full text] [doi: 10.3390/ijerph16030331] [Medline: 30691013]
106. Pratiwi PS, Tjondronegoro D. Towards personalisation of physical activity e-coach using stage-matched behaviour change and motivational interviewing strategies. In: Proceedings of the IEEE Life Sciences Conference (LSC). 2017 Presented at: IEEE Life Sciences Conference (LSC); Dec. 13-15, 2017; Sydney, NSW, Australia. [doi: 10.1109/lsc.2017.8268130]
107. Lindwall M, Ivarsson A, Weman-Josefsson K, Jonsson L, Ntoumanis N, Patrick H, et al. Stirring the motivational soup: Within-person latent profiles of motivation in exercise. Int J Behav Nutr Phys Act 2017 Jan 14;14(1):4 [FREE Full text] [doi: 10.1186/s12966-017-0464-4] [Medline: 28088208]
108. Nicklas T, Lopez S, Liu Y, Saab R, Reiher R. Motivational theater to increase consumption of vegetable dishes by preschool children. Int J Behav Nutr Phys Act 2017 Feb 07;14(1):16 [FREE Full text] [doi: 10.1186/s12966-017-0468-0] [Medline: 28166788]
109. Yu Z, Sealey-Potts C, Rodriguez J. Dietary self-monitoring in weight management: Current evidence on efficacy and adherence. J Acad Nutr Diet 2015 Dec;115(12):1931-1938. [doi: 10.1016/j.jand.2015.04.005] [Medline: 26028176]
110. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Keränen A, Salonurmi T, et al. Persuasive user experiences of a health Behavior Change Support System: A 12-month study for prevention of metabolic syndrome. Int J Med Inform 2016 Dec;96:51-61. [doi: 10.1016/j.ijmedinf.2016.02.005] [Medline: 26992482]
111. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: Mixed methods study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
112. Sharpe EE, Karasouli E, Meyer C. Examining factors of engagement with digital interventions for weight management: Rapid review. JMIR Res Protoc 2017 Oct 23;6(10):e205 [FREE Full text] [doi: 10.2196/resprot.6059] [Medline: 29061557]
113. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: A systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
114. 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]
115. 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 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
116. 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]
117. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: Implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
118. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: A systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
119. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: Application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
120. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: Can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
121. 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]
122. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13:254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
123. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]

124. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
125. Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB. Determinants of successful ehealth coaching for consumer lifestyle changes: Qualitative interview study among health care professionals. J Med Internet Res 2018 Jul 05;20(7):e237. [doi: 10.2196/jmir.9791] [Medline: 29980496]
126. Blackburn NE, Wilson JJ, McMullan II, Caserotti P, Giné-Garriga M, Wirth K, et al. The effectiveness and complexity of interventions targeting sedentary behaviour across the lifespan: A systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Apr 25;17(1):53 [FREE Full text] [doi: 10.1186/s12966-020-00957-0] [Medline: 32334631]
127. Altenburg TM, Kist-van Holthe J, Chinapaw MJ. Effectiveness of intervention strategies exclusively targeting reductions in children's sedentary time: A systematic review of the literature. Int J Behav Nutr Phys Act 2016 Jun 9;13(1):A. [doi: 10.1186/s12966-016-0387-5]
128. DeSmet A, Bastiaensens S, Van Cleemput K, Poels K, Vandebosch H, Deboutte G, et al. The efficacy of the Friendly Attac serious digital game to promote prosocial bystander behavior in cyberbullying among young adolescents: A cluster-randomized controlled trial. Comput Hum Behav 2018 Jan;78:336-347. [doi: 10.1016/j.chb.2017.10.011]
129. Nguyen P, Le LK, Nguyen D, Gao L, Dunstan DW, Moodie M. The effectiveness of sedentary behaviour interventions on sitting time and screen time in children and adults: An umbrella review of systematic reviews. Int J Behav Nutr Phys Act 2020 Sep 21;17(1):117 [FREE Full text] [doi: 10.1186/s12966-020-01009-3] [Medline: 32958052]
130. Aleksovska K, Puggina A, Giraldi L, Buck C, Burns C, Cardon G, et al. Biological determinants of physical activity across the life course: A "Determinants of Diet and Physical Activity" (DEDIPAC) umbrella systematic literature review. Sports Med Open 2019 Jan 08;5(1):2 [FREE Full text] [doi: 10.1186/s40798-018-0173-9] [Medline: 30617718]
131. Zoellner JM, Hedrick VE, You W, Chen Y, Davy BM, Porter KJ, et al. Effects of a behavioral and health literacy intervention to reduce sugar-sweetened beverages: A randomized-controlled trial. Int J Behav Nutr Phys Act 2016 Mar 22;13:38 [FREE Full text] [doi: 10.1186/s12966-016-0362-1] [Medline: 27000402]
132. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: Conceptual review and framework for intervention research. Int J Behav Nutr Phys Act 2017 Nov 15;14(1):156 [FREE Full text] [doi: 10.1186/s12966-017-0610-z] [Medline: 29141646]
133. Lloyd J, Dean S, Creanor S, Abraham C, Hillsdon M, Ryan E, et al. Intervention fidelity in the definitive cluster randomised controlled trial of the Healthy Lifestyles Programme (HeLP) trial: Findings from the process evaluation. Int J Behav Nutr Phys Act 2017 Nov 28;14(1):163 [FREE Full text] [doi: 10.1186/s12966-017-0616-6] [Medline: 29179724]
134. McKay H, Naylor P, Lau E, Gray SM, Wolfenden L, Milat A, et al. Implementation and scale-up of physical activity and behavioural nutrition interventions: An evaluation roadmap. Int J Behav Nutr Phys Act 2019 Nov 07;16(1):102 [FREE Full text] [doi: 10.1186/s12966-019-0868-4] [Medline: 31699095]
135. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: A systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
136. Mao AY, Chen C, Magana C, Caballero BK, 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]
137. Klein MC, Manzoor A, Mollee JS. Active2Gether: A personalized m-health intervention to encourage physical activity. Sensors (Basel) 2017 Jun 19;17(6):1436 [FREE Full text] [doi: 10.3390/s17061436] [Medline: 28629178]

Abbreviations

DBCI: digital behavior change intervention
DBI: digital behavior intervention
mHealth: mobile health
PA: physical activity
RQ: research question
SANRA: Scale for the Assessment of Narrative Review Articles
TUDER: Targeting, Understanding, Designing, Evaluating, and Refining

Edited by R Kukafka, G Eysenbach; submitted 04.01.21; peer-reviewed by M Savolainen, F Tzelepis, K Ross; comments to author 16.03.21; revised version received 07.05.21; accepted 10.08.21; published 17.11.21.

Please cite as:
Chatterjee A, Prinz A, Gerdes M, Martinez S
Digital Interventions on Healthy Lifestyle Management: Systematic Review
J Med Internet Res 2021;23(11):e26931
URL: https://www.jmir.org/2021/11/e26931
doi:10.2196/26931
PMID:34787575

Review

Effectiveness of Information and Communication Technology on Obesity in Childhood and Adolescence: Systematic Review and Meta-analysis

Jihyun Park[1], MD; Mi-Jeong Park[1], MD; Young-Gyun Seo[1], MD, PhD

Department of Family Medicine, Hallym University Sacred Heart Hospital, Anyang, Republic of Korea

Corresponding Author:
Young-Gyun Seo, MD, PhD
Department of Family Medicine
Hallym University Sacred Heart Hospital
22, Gwanpyeong-ro 170beon-gil
Dongan-gu
Anyang, 14068
Republic of Korea
Phone: 82 31 380 3805
Fax: 82 31 380 1782
Email: yg035@daum.net

Abstract

Background: Internet or mobile device use as a form of information and communication technology (ICT) can be more effective in weight loss and weight maintenance than traditional obesity interventions.

Objective: The study aims to assess the effectiveness of child-centered ICT interventions on obesity-related outcomes.

Methods: Articles were retrieved from the Cochrane Central Register of Controlled Trials, Embase, and PubMed web-based databases. We selected randomized controlled trials in which the participants were aged <18 years. The primary outcomes were BMI, body weight, BMI *z*-score, waist circumference, and percentage body fat.

Results: In total, 10 of the initial 14,867 studies identified in the databases were selected according to the inclusion criteria. A total of 640 participants were included in the intervention group and 619 in the comparator group. Meta-analyses were conducted considering various subgroups (intervention type, comparator type, target participants, mean age, sex, BMI status, and follow-up period). Overall, ICT interventions demonstrated no significant effect on BMI, body weight, BMI *z*-score, waist circumference, and percentage body fat. Subgroup analyses revealed that the effect of the intervention was statistically significant for the following: *web* intervention (weighted mean difference [WMD]=−1.26 kg/m^2, 95% CI −2.24 to −0.28), *lifestyle modification* comparator (WMD=−1.75, 95% CI −2.76 to −0.74), intervention involving both *boys and girls* (WMD=−1.30, 95% CI −2.14 to −0.46), and intervention involving *obesity only* (WMD=−1.92, 95% CI −3.75 to −0.09).

Conclusions: The meta-analysis results for children with obesity who used the *web* intervention program confirmed significant effects on BMI reduction compared with *lifestyle modification*. Evidence from the meta-analysis identified internet technology as a useful tool for weight loss in children with obesity.

(J Med Internet Res 2021;23(11):e29003) doi:10.2196/29003

KEYWORDS

ICT; eHealth; mHealth; weight loss; obesity; BMI; meta-analysis; randomized controlled trial; children; adolescents; mobile phone

Introduction

Background

The prevalence of obesity among children and adolescents worldwide has increased at an alarming rate [1,2]. Children with obesity are now prone to developing diseases that were only observed in adults, including high blood pressure, impaired glucose tolerance, type 2 diabetes mellitus, dyslipidemia, sleep apnea, joint problems, and fatty liver disease [3-5]. In addition, children with obesity may be at risk for a variety of social and

psychological problems such as low self-esteem, bullying, and discrimination [6].

Given the adverse health outcomes and high prevalence rate, effective interventions are imperative for the management of childhood obesity. Traditionally, many forms of intervention have been attempted to address childhood obesity, and they have mainly comprised lifestyle modifications and mental health care as well as medication and surgical treatment [7-10]. However, some methods have limited indications in children, and their effectiveness is debatable [11,12]. Therefore, it is necessary to develop more efficient and effective approaches for children.

Information and Communication Technology

Internet and mobile use in childhood and adolescence are already becoming essential elements of young people's lives [13-15], providing several advantages such as learning, information, and entertainment but also causing many problems. Excessive internet and mobile use has resulted in more sedentary behavior, decreased physical activity, and unhealthy dietary patterns, and it is emerging as a social problem that suffices diagnosis as a form of addiction disorder [16]. In addition, studies have reported that increased screen time and obesity are strongly correlated [17].

In contrast, there have been some attempts to use active internet and mobile use during childhood as a means of obesity intervention. Mobile health (mHealth), defined as *medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices* [18], has the potential to influence a variety of health outcomes and has become a key trend in health service provision during recent years [19]. In addition, previous research has demonstrated that internet-based behavioral interventions have the potential for weight management [20-22].

Information and communication technology (ICT) is a technology that can be used to connect information technologies such as computers and software with communication technologies such as telephones and telecommunication networks. ICT includes cell phone calls, SMS, and apps using a mobile phone, email, and web services using a computer, and telehealth, including health education services, remote monitoring, and remote counseling [23].

Objectives

Internet or mobile device use as a form of ICT can be more effective in weight loss and weight maintenance than traditional obesity intervention, as it offers benefits in terms of cost, ease of use, accessibility, and time to visit, while improving compliance with prescribed treatments through extensive patient monitoring and continuous support.

Therefore, we conducted a systematic review and meta-analysis to assess the effectiveness of child- and adolescent-centered ICT interventions on obesity-related outcomes.

Methods

Overview

We performed a meta-analysis based on the Cochrane Handbook for Systematic Reviews of Interventions [24] and the Centre for Reviews and Dissemination's guidance for undertaking reviews in health care [25]. We reported based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [26].

Literature Search

A systematic search of the effects of ICT on obesity-related outcomes was conducted. We searched the Cochrane Central Register of Controlled Trials, Embase, and PubMed web-based databases and retrieved articles published before January 1, 2021. The search terms used were as follows: *ICT OR information and communication technology OR Internet OR web OR social media OR mobile OR smartphone OR application OR app* AND *obesity OR obese OR weight OR metabolic syndrome*. The search was limited to randomized controlled trials (RCTs) and English articles; however, there were no restrictions on the calendar date. Reference lists of the retrieved articles were also reviewed. Information that was unavailable in the selected articles was requested by contacting the relevant authors; however, no response was received.

Two of the authors (JP and MJP) independently reviewed the titles and abstracts after the removal of duplicates. Discrepancies were resolved either by a discussion between the authors or by requesting comments from the third author (YGS). The 3 authors independently analyzed the full text of the remaining articles to determine the final inclusion.

Eligibility Criteria

We selected the trials to be included in the meta-analysis using the following criteria: (1) the trial was a human RCT written in English, and the full text was available; (2) participants were aged <18 years; (3) the intervention group underwent ICT intervention alone or along with other lifestyle interventions; (4) the comparator group did not undergo ICT intervention; (5) the trial included an assessment of the following primary outcomes: BMI, body weight (BW), BMI *z*-score, waist circumference (WC), and percentage body fat (%BF); and (6) mean values of changes from baseline (or postintervention values if not available) with SD (or data suitable for calculating SD: 95% CI or SE). Uncontrolled, cross-sectional, and animal studies were excluded. The selection criteria did not limit the type of ICT used.

Risk of Bias Assessment

We used guidelines from the Cochrane Handbook for Systematic Reviews of Interventions to assess the risk of bias in the RCTs [24]. Sources of bias, such as selection bias (random sequence generation and allocation concealment), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), and reporting bias (selective reporting) were evaluated. Each domain was assessed in terms of methodological quality, with low or high risk of bias. If data were insufficient to make

a reasonable judgment, the domain was described as *unclear risk of bias*.

The risk of bias was reported graphically using Review Manager (RevMan, Version 5.3; Copenhagen: The Nordic Cochrane Center, The Cochrane Collaboration, 2014).

Data Extraction

Data were independently extracted by 2 authors (JP and YGS) from the selected RCTs. From each RCT, the following data were extracted: name of the first author, year of publication, country where the RCT was performed, sample size, participant-related variables (age, sex, and BMI status), intervention-related variables (ICT type, study duration, target participants, intervention details, comparator details, intervention frequency, and feedback frequency), and treatment effects (mean difference and SD of 2 time point values or mean and SD of postintervention values). The primary outcomes were BMI, BW, BMI *z*-score, WC, and %BF.

Data Synthesis

The data set was constructed using the mean differences and SDs between the pre- and postintervention values. When the mean difference and SD were not published, the mean and SD of the postintervention values were used. In a meta-analysis, it was possible to combine both the mean differences and the means of postintervention values, assuming that the relative effects assessed by both the mean differences and the means of postintervention values are the same [24]. The final results were calculated and aggregated by one author (YGS).

Meta-analysis

For the meta-analysis, we used Stata/MP (version 14.0; StataCorp). The weighted mean differences (WMDs) of BMI, BW, BMI *z*-score, WC, and %BF in the intervention and comparator groups were calculated. We used the Cochran Q test and I^2 test to test the heterogeneity between the study results. For interpretation, I^2 values of 25, 50, and 75 were considered to represent low, moderate, and high heterogeneity, respectively [27]. To consider heterogeneity, the DerSimonian and Laird [28] random-effects model for estimating WMD with 95% CI was used. The effect size and 95% CI of each study were expressed as forest plots. We checked the symmetry of the funnel plots to evaluate the presence of publication bias. In addition, we used the Egger regression test to evaluate the small study effects [29]. Heterogeneity between studies was analyzed using a meta-regression. We used covariates that may influence the association between ICT and BMI, namely, intervention type (web vs web plus vs app vs app plus), comparator type (control vs print-based vs lifestyle modification), target participants (parents and children vs children vs parents), mean age (<10 vs ≥10 years), sex (boys and girls vs boys vs girls), BMI status (normal to obese vs overweight or obesity vs obesity), and follow-up period (≥6 vs <6 months). The cutoff for the intervention period (6 months) was based on the transtheoretical model [30]. The statistical significance level was set at 5%. For heterogeneity, a threshold *P* value of .10 estimated using the Cochran Q test was considered statistically significant [27].

Results

Study Selection and Characteristics

In total, 10 [31-40] of the initial 14,867 studies identified in the databases were selected according to the inclusion criteria, and they contained sufficient data for meta-analysis (Figure 1). The meta-analysis included 13 data sets (2 studies each had 2 ICT types [32,34]: web and web plus; one study had two comparator types [38]: lifestyle modification and lifestyle modification plus). A total of 640 participants were included in the intervention group (range of the number of participants, 15-181) and 619 (range of the number of participants, 13-180) in the comparator group. All participants were aged <18 years. Of the included studies, one evaluated boys only [40], another evaluated girls only [39], and the other 8 did not differentiate between the sexes of the participants. Six out of 10 studies had intervention periods of 12 weeks, and the remaining 4 had intervention periods >12 weeks. Five studies had follow-up periods of 12 weeks, and the remaining 5 had follow-up periods of >12 weeks. The frequency of interventions and feedback varied from study to study. The average attendance rate of the participants in the study was 86.49% (1089/1259). The characteristics of the selected RCTs are summarized in Table 1.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection. ICT: information and communication technology; RCT: randomized controlled trial.

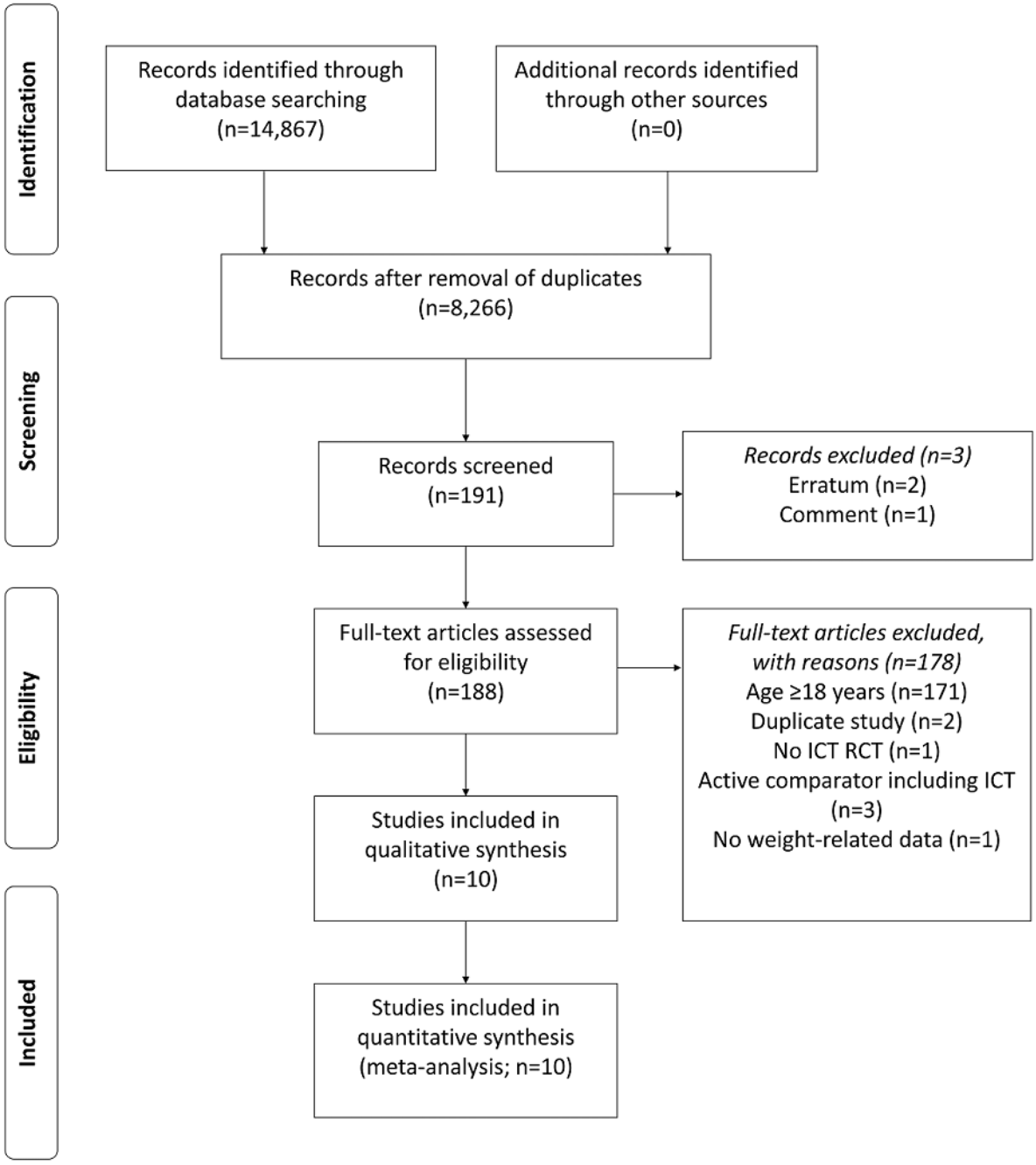


Table 1. Characteristics of the selected randomized clinical trials.

Study (country)	Number at baseline→follow-up	Age, range or mean (SD)	Sex (% males)	BMI, range or mean (SD)	BMI status	ICT[a] type	Intervention duration (weeks)	Target participants	Intervention details	Comparator details	Intervention frequency	Feedback frequency
Johansson et al [31] (Sweden)	IG[b] 15→9; CG[c] 13→9	5-12	Male or female (46.4)	Boys: BMI ≥98.9th percentile; girls: BMI ≥98.6th percentile	Obesity	App	24	Parent and children	Provement: display weight loss target curve, exchange text messages; Lifee Spirits+activity monitor: increase motivation for physical activity	LSM[d] (improving dietary habits and increasing physical activity to reduce the degree of obesity)	Daily monitoring, messages whenever they felt a need for support	At least weekly
Chai et al [32] (Australia)	IG (telehealth) 16→11; IG (telehealth+SMS) 15→10; CG 15→15	4-11; 9 (2.3)	Male or female (59)	22.5 (5.1)	Overweight or obesity	Telehealth or telehealth+SMS	12	Parent and children	Telehealth dietitian consultation: semistructured telehealth consultations (approximately 20 minutes each); website (Back2Basics Family): information on various nutrition topic; Facebook group: exchange ideas and information related to the Back2Basics Family website; SMS to parents: targeting healthy eating for children	Waitlist control	Telehealth: week 1, 4; website: preferred time and frequency; Facebook: weekly; SMS: 4-weekly rotations of decreasing frequency (ie, 5, 4, 3, 2 per week)	Week 1, 4

Study (country)	Number at baseline→follow-up	Age, range or mean (SD)	Sex (% males)	BMI, range or mean (SD)	BMI status	ICT[a] type	Intervention duration (weeks)	Target participants	Intervention details	Comparator details	Intervention frequency	Feedback frequency
Delisle Nyström et al [33] (Sweden)	IG 156→133; CG 159→130	4.5	Male or female (39)	IG: 15.9 (1.4); CG: 15.7 (1.2)	Normal to obesity	App	24	Parent	General information, advice, and evidence-based strategies on how to change unhealthy behaviors; register child's intake of fruits, vegetables, candy, sweetened beverages, and sedentary time; submit questions to a dietician and a psychologist to ask questions specific to their child	A pamphlet on healthy eating and physical activity	At least weekly	Weekly
Bruñó et al [34] (Spain)	IG (Move It) 18→15; IG (Move It plus) 16→15; CG 18→13	9-16; 12.6 (1.7)	Male or female (57.7)	BMI ≥85th percentile	Overweight or obesity	Web or web+email	12	Children	Move It: web-based physical exercise program combines one aerobic exercise (brisk walking) and 10 muscular strength exercises; Move It plus: move It+weekly reminder and motivational emails	The same exercise program as the intervention group by a written guide	60 sessions distributed over 3 months, with 5 weekly sessions of 60 minutes each	Weekly

Study (country)	Number at baseline→follow-up	Age, range or mean (SD)	Sex (% males)	BMI, range or mean (SD)	BMI status	ICT[a] type	Intervention duration (weeks)	Target participants	Intervention details	Comparator details	Intervention frequency	Feedback frequency
Rerksuppaphol and Rerksuppaphol [35] (Thailand)	IG 111→111; CG 107→106	10.7 (3.1)	Male or female (49)	8.36 kg/m^2 (IQR 16.08-22.09 kg/m^2)	Normal to obesity	Web	16	Children	Personal data collection, anthropometric variables and the interpretation of nutritional status, information related to healthy nutrition, food habits and physical activity by web	The same program as the intervention group by trained research assistants	Monthly	Monthly
Mameli et al [36] (Italy)	IG 23→16; CG 20→14	10-17	Male or female (61.9)	BMI ≥95th percentile	Obesity	App+wrist band	12	Children	App: measure energy intake; wrist band: measure energy expenditure; SMS: feedback	LSM (the Mediterranean diet and instruction to practice physical activity and minimize sedentary activity)	Data obtained by the wrist band and app were made available daily	Weekly
Mohammed Nawi and Che Jamaludin [37] (Malaysia)	IG 47→47; CG 50→50	16	Male or female (56.7)	BMI >25 kg/m^2	Overweight or obesity	Web	12	Children	Information on healthy lifestyle, diet, and ways to overcome obesity, discussion by web	The same information as the intervention group by the pamphlets	Weigh and calculate BMI every 2 weeks; notified with any updates and information	Chat sessions on the website; monitoring negative comment by admin
Abraham et al [38] (China)	IG 16→16; CG (sLMP[e]) 16→16; CG (control) 16→16	12-18	Male or female (60.4)	BMI ≥95th percentile	Obesity	Web+cell phone calls+SMS	12	Children	Internet-based curriculum (nutrition and physical activity); cell phone follow-up; weekly semipersonalized SMS	Control: usual care consisted of a focused dietary and physical activity history, medical history, physical examination, laboratory screening and obesity counseling; sLMP: usual care+four meetings with a nutritionist over 3 months	Goal setting: monthly	Weekly SMS

Study (country)	Number at baseline→follow-up	Age, range or mean (SD)	Sex (% males)	BMI, range or mean (SD)	BMI status	ICT[a] type	Intervention duration (weeks)	Target participants	Intervention details	Comparator details	Intervention frequency	Feedback frequency
Nollen et al [39] (United States)	IG 26→22; CG 25→22	9-14; 11.3 (1.6)	Female (0)	23.7 (5.7)	Normal to obesity	Standalone mobile app	12	Children	Set 2 daily goals and an accompanying plan for improving the behavior, and feedback and reinforcement on goal-attainment	The same contents as the intervention group by the manuals	Self-monitor progress toward their goals at 5 times a day	Five times a day
Smith et al [40] (Australia)	IG 181→139; CG 180→154	12-14; 12.7 (0.5)	Male (100)	20.5 (4.1)	Normal to obesity	App	20	Children	Supplement the delivery of enhanced school sport and interactive sessions by providing participants with physical activity monitoring, recording of fitness challenge results, tailored motivational messaging, peer assessment of resistance training skills, and goal setting for physical activity and screen-time	Participate in usual practice (regularly scheduled school sports and physical education lessons)	Tailored motivational and informational *push prompt* messages	Peer assessment

[a]ICT: information and communication technology.

[b]IG: intervention group.

[c]CG: comparator group.

[d]LSM: lifestyle modification.

[e]sLMP: simplified lifestyle modification program.

ICT Interventions

Of the included 10 studies, 5 (50%) were web-based interventions, and the other 5 (50%) were app-based interventions.

Two of the 5 web-based intervention studies were *web* interventions, one for providing web-based nutrition and physical activity information for children [35] and one for providing web-based nutrition information for children [37]. One of the 5 web-based intervention studies was *web* or *web plus* intervention, which provided web-based physical exercise programs and motivational emails for children [34]. Two of the 5 web-based intervention studies were *web plus* interventions, one for providing web-based nutrition and physical activity information as well as SMS feedback for children [38] and one for providing web-based telehealth dietitian consultation, nutrition information, and SMS feedback for parents and children [32].

Four of the 5 app-based intervention studies were *app* interventions, one for providing app-based weight loss target curve, physical activity information for parents and children [31], one for providing app-based nutrition information for parents [33], one for providing app-based nutrition and screen-time information for children [39], and one for providing app-based physical activity and screen-time information for children [40]. One of the 5 app-based intervention studies was *app plus* intervention, which provided app-based nutrition and physical activity information and SMS feedback for children [36].

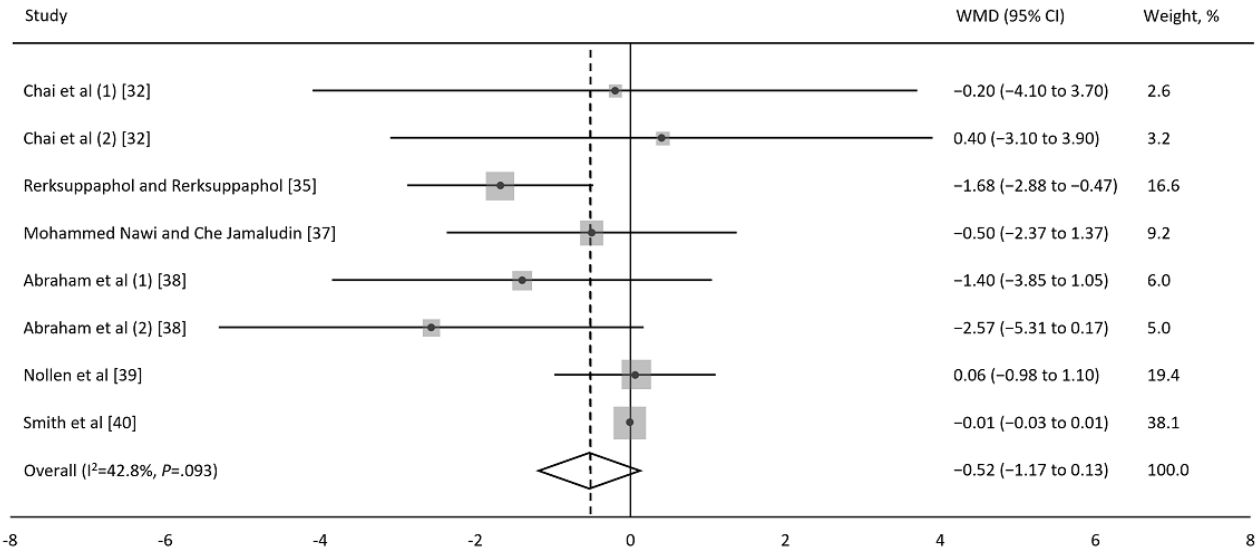
Risk of Bias

Participant blinding was not possible because of the characteristics of the intervention. Therefore, performance bias was not considered in the risk of bias assessment. There was some risk of bias in the individual studies. Two studies lacked sufficient data to evaluate the randomization sequence generation. Four studies lacked sufficient data to evaluate allocation concealment. Three studies indicated blinding of outcome assessment; however, one study stated that assessors were not blinded at follow-up. Two studies lacked sufficient data to evaluate the attrition bias. Three studies lacked sufficient information to evaluate the study protocol, and one study did not report all of the information. The risk of bias assessment is reported graphically in Figure S1 in Multimedia Appendix 1 [31-40].

Synthesis of Results

To evaluate the overall intervention effect, we calculated the mean difference in BMI for each study. Figure 2 shows the effect size for each study and the overall effect size. The intervention demonstrated no significant effect on BMI (WMD=−0.52 kg/m^2, 95% CI −1.17 to 0.13). Categorization of target participants, mean age, and BMI appeared to have moderate to high heterogeneity. Categorization of the follow-up period appeared to have low to moderate heterogeneity. For all other categories, the heterogeneity was low. We also calculated mean differences in BW, BMI *z*-score, WC, and %BF to determine the overall intervention effect, and no significant intervention effects were identified for BW (WMD=−0.22 kg, 95% CI −1.05 to 0.62), BMI *z*-score (WMD=−0.22, 95% CI −0.49 to 0.04), WC (WMD=−1.70 cm, 95% CI −3.91 to 0.51), and %BF (WMD=−0.00%, 95% CI −0.07 to 0.07; Figure S2 in Multimedia Appendix 1).

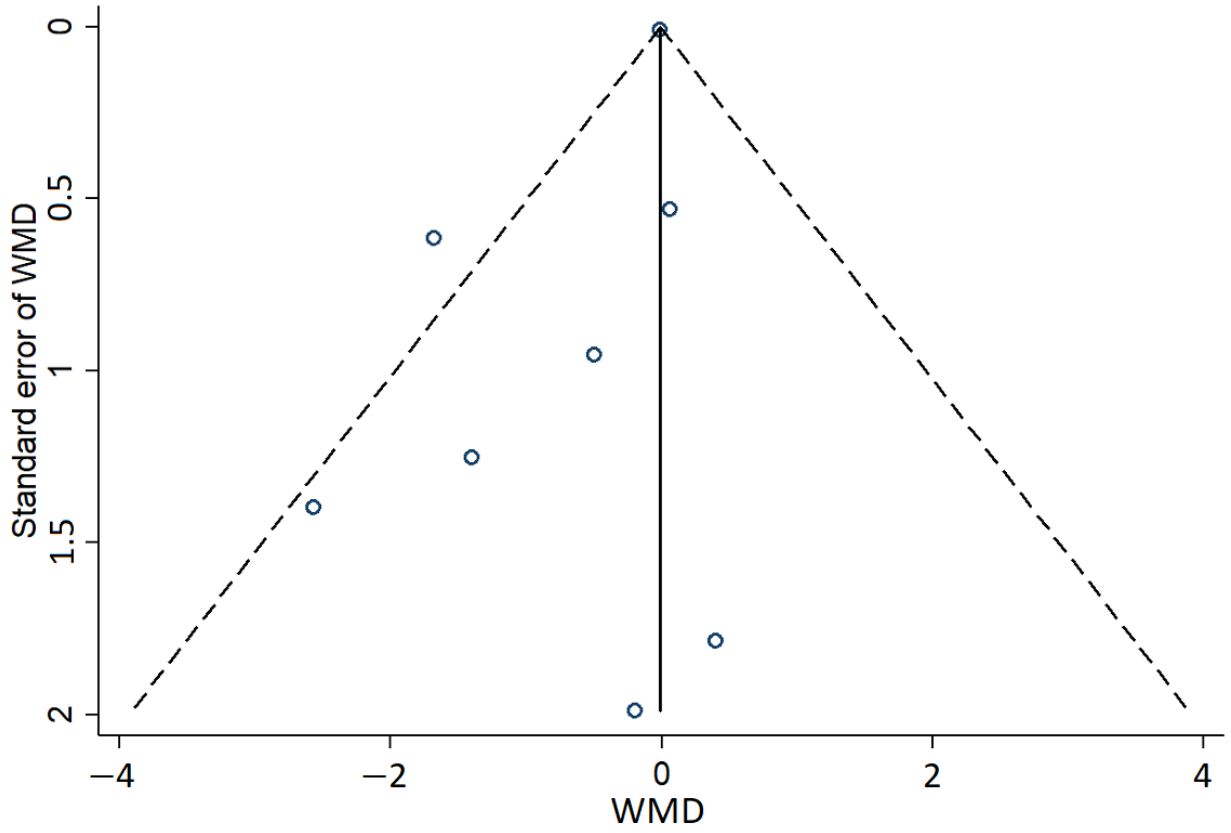
Figure 2. Forest plot for changes in BMI. Meta-analysis of the effect of the information and communication technology (ICT) on BMI (kg/m^2). The mean difference for each study reporting changes in BMI is depicted along with the 95% CI. The random-effects model was used to estimate the weighted mean differences with 95% CIs. Negative values favor ICT because the ICT intervention group experienced more BMI reduction than the comparator group did. WMD: weighted mean difference.



Publication Bias

To verify possible publication bias, we plotted the effect size against the SE to generate a funnel plot (Figure 3). There was no statistically significant publication bias according to the Egger test (*P*=.09).

Figure 3. Funnel plot for changes in BMI. The funnel plots of SE of weighted mean difference (WMD) against WMD for BMI to assess for publication bias. WMD: weighted mean difference.



Meta-regression

The results of the simple meta-regression analysis were significant for categorical covariates of intervention type (β=.69, 95% CI 0.10 to 1.28) and comparator type (β=−.73, 95% CI −1.39 to −0.07). A subgroup analysis by intervention type revealed that the intervention effect was statistically significant only in the *web* intervention (WMD=−1.26 kg/m^2, 95% CI −2.24 to −0.28). A subgroup analysis by comparator type demonstrated that the intervention effect was statistically significant only in the *lifestyle modification* comparator (WMD=−1.75 kg/m^2, 95% CI −2.76 to −0.74). A subgroup analysis by sex revealed that the intervention effect was statistically significant only in the intervention involving both *boys and girls* (WMD=−1.30 kg/m^2, 95% CI −2.14 to −0.46). In addition, a subgroup analysis by BMI status demonstrated that the intervention effect was statistically significant only in the intervention involving *obesity only* (WMD=−1.92 kg/m^2, 95% CI −3.75 to −0.09; Figure S3 in Multimedia Appendix 1).

Subgroup analyses were also performed for other outcome variables, and the results of subgroups containing only one study data were excluded. For BW, subgroup analysis revealed that the intervention effect was statistically significant only in the *web* intervention (WMD=−1.21 kg, 95% CI −2.36 to −0.06) and the intervention involving a follow-up period *<6 months* (WMD=−0.87 kg, 95% CI −1.73 to −0.01; Figure S4 in Multimedia Appendix 1). For BMI *z*-score, subgroup analysis demonstrated that the intervention effect was statistically significant only in the intervention targeting *parents and children* (WMD=−0.23, 95% CI −0.37 to −0.08) and that where the average age of participants was *<10 years* (WMD=−0.23, 95% CI −0.37 to −0.08; Figure S5 in Multimedia Appendix 1). For WC, subgroup analysis revealed that the intervention effect was statistically significant only in the *web plus* intervention (WMD=−4.88 cm, 95% CI −8.93 to −0.83), *lifestyle modification* and *control* comparator (WMD=−3.58 cm, 95% CI −6.30 to −0.85 and WMD=0.50 cm, 95% CI 0.43 to 0.57, respectively), intervention involving both *boys and girls* (WMD=−2.95 cm, 95% CI −5.18 to −0.71), and intervention involving *obesity only* (WMD=−5.50 cm, 95% CI −9.89 to −1.12; Figure S6 in Multimedia Appendix 1). Subgroup analyses demonstrated no statistically significant effect on %BF (Figure S7 in Multimedia Appendix 1).

Behavioral Change

Nine of the included 10 studies reported the initial results, and only one study reported the results of an additional 6 months follow-up after reporting the initial results of a 6-month intervention [33].

Five of the included 10 studies also reported results on behavioral changes. Some studies have shown that there were no significant intervention effects on sedentary time [33], physical activity [33,38], consumption of fruits and vegetables,

and sugar-sweetened beverages [33,39]. However, one study reported that screen-time and sugar-sweetened beverage consumption [40] were improved, and another study reported that dietary intake was improved in the intervention group [32]. One study reported no significant intervention effects on stress or dietary knowledge scores [38]. However, another study reported that emotional functioning significantly increased after the intervention [37].

Discussion

Principal Findings

This study identified the effects of ICT intervention on obesity-related outcomes in children and adolescents in RCTs through a systematic literature review and meta-analysis. We found that although ICT intervention is not significantly effective in reducing BMI, BW, BMI z-score, WC, and %BF, it does work in certain groups.

A subgroup analysis by intervention type revealed that the intervention effect was statistically significant only in the *web* or *web plus* intervention for BMI, BW, and WC. This result is consistent with those of previous studies that showed that web-based health programs are effective in managing obesity. Meta-analyses have demonstrated that web-based interventions are effective in achieving weight loss in adults [41,42]. Several systematic reviews have explored the use of web-based interventions for the prevention or treatment of obesity and related conditions in pediatric populations [20,22]. In addition, a meta-analysis demonstrated that mobile-based interventions are effective in achieving weight loss [43,44] and reducing BMI [45-47] in adults. However, there has been no meta-analysis of mobile-based obesity intervention studies in children. Several systematic reviews have indicated that mobile-based interventions in obesity treatment programs have a modest effect on weight control [47-51]. However, these effects are inconsistent. Owing to the nature of mobile use, access is possible from anywhere; hence, the possibility of giving formal responses in situations where participants are unprepared to take certain actions cannot be ruled out. Therefore, additional RCTs and meta-analyses targeting the weight-loss effect of mobile-based interventions in children and adolescents are warranted.

Another subgroup analysis by comparator type revealed that the intervention effect was significant only in the *lifestyle modification* comparator rather than in the *control* comparator for BMI and WC. A previous meta-analysis [52] that analyzed the weight loss effect of circuit training demonstrated that focusing on circuit training alone rather than adding other lifestyle interventions to circuit training is effective. Among the RCTs included in our meta-analysis, 2 [35,38] had the *lifestyle modification* comparator. In these 2 studies, the comparator group did not focus on one lifestyle, intervened in various lifestyle habits, and the number of contacts was not frequent. Recognizing that they are undergoing an intervention, participants tend to act passively, hoping to elicit changes to their lifestyle habits; therefore, where there is no intensive intervention, the intervention can be counterproductive. However, as there may be other factors influencing the research results, additional studies comparing ICT with a comparator group focusing on one lifestyle are needed.

For WC, a significant intervention effect was also obtained when the comparator type was *control*. Among the RCTs included in our meta-analysis, 2 [32,40] had the *control* comparator. However, the difference in sample size was large between the 2 studies; thus, the results of the subgroup were not different from those of one study [40], which had a large sample size. Therefore, it is unreasonable to interpret this as a subgroup result.

The intervention effect was significant only in the intervention involving both *boys and girls* for BMI and WC. Among the RCTs included in our meta-analysis, one evaluated boys only [40], another evaluated girls only [39], and the other 8 did not differentiate between the sexes of the participants. Sex differences in response to ICT interventions are likely to be because of differences in participation. In cases where boys and girls participate together, the resultant mutual competition can increase participation, which subsequently increases the effect of the intervention. Several studies have compared the sex-dependent effects of school-based physical activity interventions [53-56]. However, to date, no meta-analysis has shown sex differences in the effects of ICT intervention on children with obesity. Therefore, additional RCTs are needed to investigate the sex-dependent weight loss effects of ICT interventions.

In addition, the intervention effect was significant only in the intervention involving obesity only for BMI and WC. Participants with normal weight or overweight status may be less motivated than with participants with obesity, which may dilute the overall effect. It is encouraging to identify meaningful results from studies solely involving children and adolescents with obesity. In other words, significant results can be obtained if ICT intervention is implemented in children and adolescents with obesity. Further well-designed ICT intervention studies targeting children and adolescents with obesity should be conducted.

When both parents and children were involved in children less than 10 years of age, a reduction in BMI *z*-score was observed. In the same situation, other outcome variables did not show a statistically significant effect; however, considering the BMI *z*-score only, ICT could be used as a means of preventing obesity-related outcomes by intervening in early childhood with the involvement of parents. Therefore, there is a need for additional long-term RCTs involving high-quality ICT interventions targeting parents and children under 10 years of age.

Previous studies have demonstrated that mHealth programs or internet technology have a higher attrition rate than conventional face-to-face methods [57,58]. However, the average attendance rate of the participants in this meta-analysis was 86.5%. Children with obesity are often ashamed because of social stigma and may face prejudice against weight [59]. Moreover, because children and teenagers have a high level of interest in and concentration with electronic devices, access through mobile means or the internet can be more efficient [60]. Therefore, the effect of obesity intervention through mHealth programs or

internet technology is considered superior because of enhanced accessibility and an increased participation rate compared with conventional interventions such as the face-to-face method.

The advantage of childhood-obesity management using mHealth programs or internet technology is that it is possible to operate programs led by peer participants and provide real-time feedback. In particular, by presenting a mission aimed at improving dietary habits or increasing exercise among peers as well as providing rankings or rewards, it is possible to induce mutual participation through goodwill competition. However, among the studies included in this study, no study was conducted in a manner that induced competition among participants. Therefore, further research on the participation rates is imperative.

Strengths

Our meta-analysis had the following strengths: First, to the best of our knowledge, this is the first meta-analysis to assess the association between ICT intervention and weight loss in children and adolescents. Second, we examined the differences in weight-loss effects among subgroups according to the type of intervention, type of comparator, target participants, mean age, sex, BMI status, and follow-up period. We found that ICT intervention is effective for weight loss in the *web* intervention, *lifestyle modification* comparator, intervention involving both *boys and girls*, and intervention involving *obesity only*. Finally, the heterogeneity among the included RCTs was low.

Limitations

Our meta-analysis has several limitations. Although the meta-analysis found a moderate effect size that was statistically significant, as not many studies were included in the meta-analysis, generalization of the study results is limited because of potential publication bias. However, there were no small study effects according to the Egger regression test results. In addition, many of the studies were of short duration, making it unclear whether weight loss was sustained in the long term. Although the clinically significant threshold for weight loss was not always achieved across the studies, studies of longer duration might have found clinically significant weight loss. Therefore, further evidence is necessary to confirm this hypothesis. Outcome measurements based on participants' self-reports for many studies were recorded using mobile apps and websites and the lack of comments on the reliability of the measurement method using mobile apps can limit the results analysis. In addition, no RCT has focused on sex differences in the use of mHealth programs and internet technology. Therefore, future research should investigate the sex-dependent weight loss effects of ICT interventions.

Conclusions

The meta-analysis results for children and adolescents with obesity who participated in the *web* intervention program confirmed significant effects on BMI reduction compared with the *lifestyle modification* intervention. Evidence from the meta-analysis identified internet technology as a useful tool for weight loss in children and adolescents with obesity.

Authors' Contributions

YGS and JP wrote the manuscript; YGS, JP, and MJP analyzed the data and interpreted the results; YGS participated in the design of the study; all authors reviewed the manuscript, contributed to the discussion, and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Risk of bias assessment and forest plots.
[DOCX File , 223 KB - jmir_v23i11e29003_app1.docx]

References

1. Fact sheets: obesity and overweight. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-03-01]
2. Ogden CL, Carroll MD, Lawman HG, Fryar CD, Kruszon-Moran D, Kit BK, et al. Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. J Am Med Assoc 2016 Jun 07;315(21):2292-2299. [doi: 10.1001/jama.2016.6361] [Medline: 27272581]
3. Tirosh A, Shai I, Afek A, Dubnov-Raz G, Ayalon N, Gordon B, et al. Adolescent BMI trajectory and risk of diabetes versus coronary disease. N Engl J Med 2011 Apr 07;364(14):1315-1325. [doi: 10.1056/nejmoa1006992]
4. Skinner AC, Perrin EM, Moss LA, Skelton JA. Cardiometabolic risks and severity of obesity in children and young adults. N Engl J Med 2015 Oct;373(14):1307-1317. [doi: 10.1056/NEJMoa1502821] [Medline: 26422721]
5. Twig G, Yaniv G, Levine H, Leiba A, Goldberger N, Derazne E, et al. Body-Mass Index in 2.3 million adolescents and cardiovascular death in adulthood. N Engl J Med 2016 Jun 23;374(25):2430-2440. [doi: 10.1056/nejmoa1503840]
6. Mühlig Y, Antel J, Föcker M, Hebebrand J. Are bidirectional associations of obesity and depression already apparent in childhood and adolescence as based on high-quality studies? A systematic review. Obes Rev 2016 Mar 18;17(3):235-249. [doi: 10.1111/obr.12357] [Medline: 26681065]

7. Stice E, Shaw H, Marti CN. A meta-analytic review of obesity prevention programs for children and adolescents: the skinny on interventions that work. Psychol Bull 2006 Sep;132(5):667-691 [FREE Full text] [doi: 10.1037/0033-2909.132.5.667] [Medline: 16910747]
8. Haynos AF, O'Donohue WT. Universal childhood and adolescent obesity prevention programs: review and critical analysis. Clin Psychol Rev 2012 Jul;32(5):383-399. [doi: 10.1016/j.cpr.2011.09.006] [Medline: 22681912]
9. Seo DC, Sa J. A meta-analysis of obesity interventions among U.S. minority children. J Adolesc Health 2010 Apr;46(4):309-323. [doi: 10.1016/j.jadohealth.2009.11.202] [Medline: 20307819]
10. Luttikhuis HO, Baur L, Jansen H, Shrewsbury VA, O'Malley C, Stolk RP, et al. Interventions for treating obesity in children. Cochrane Database Syst Rev 2009(1):CD001872. [doi: 10.1002/14651858.CD001872.pub2] [Medline: 19160202]
11. Kelly AS, Fox CK, Rudser KD, Gross AC, Ryder JR. Pediatric obesity pharmacotherapy: current state of the field, review of the literature and clinical trial considerations. Int J Obes (Lond) 2016 Jul 26;40(7):1043-1050 [FREE Full text] [doi: 10.1038/ijo.2016.69] [Medline: 27113643]
12. Inge TH, Courcoulas AP, Jenkins TM, Michalsky MP, Helmrath MA, Brandt ML, Teen-LABS Consortium. Weight loss and health status 3 years after bariatric surgery in adolescents. N Engl J Med 2016 Jan 14;374(2):113-123 [FREE Full text] [doi: 10.1056/NEJMoa1506699] [Medline: 26544725]
13. Lenhart A. Teens, social media and technology overview 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2021-03-01]
14. Mascheroni G, Olafsson K, Cuman A, Dinh T, Haddon L, Jørgensen H, et al. Mobile internet access and use among European children: initial findings of the Net Children Go Mobile project. In: Net Children Go Mobile Initial Findings Report. Milan, Italy: Educatt; 2013.
15. Roberts DF, Foehr UG, Rideout V. Generation M: Media in the lives of 8-18 year-olds. Kaiser Family Foundation. 2005. URL: https://www.kff.org/wp-content/uploads/2013/01/generation-m-media-in-the-lives-of-8-18-year-olds-report.pdf [accessed 2021-03-01]
16. Brenner V. Psychology of computer use: XLVII. Parameters of Internet use, abuse and addiction: the first 90 days of the Internet Usage Survey. Psychol Rep 1997 Jun 01;80(3 Pt 1):879-882. [doi: 10.2466/pr0.1997.80.3.879] [Medline: 9198388]
17. Fang K, Mu M, Liu K, He Y. Screen time and childhood overweight/obesity: a systematic review and meta-analysis. Child Care Health Dev 2019 Sep;45(5):744-753. [doi: 10.1111/cch.12701] [Medline: 31270831]
18. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Geneva: World Health Organization; 2020:1-111.
19. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
20. Nguyen B, Kornman KP, Baur LA. A review of electronic interventions for prevention and treatment of overweight and obesity in young people. Obes Rev 2011 May;12(5):298-314. [doi: 10.1111/j.1467-789X.2010.00830.x] [Medline: 21348921]
21. Antwi F, Fazylova N, Garcon MC, Lopez L, Rubiano R, Slyer JT. The effectiveness of web-based programs on the reduction of childhood obesity in school-aged children: a systematic review. JBI Libr Syst Rev 2012;10(42 Suppl):1-14. [doi: 10.11124/jbisrir-2012-248] [Medline: 27820152]
22. An JY, Hayman LL, Park YS, Dusaj TK, Ayres CG. Web-based weight management programs for children and adolescents: a systematic review of randomized controlled trial studies. ANS Adv Nurs Sci 2009;32(3):222-240. [doi: 10.1097/ANS.0b013e3181b0d6ef] [Medline: 19707091]
23. Casanova G, Zaccaria D, Rolandi E, Guaita A. The effect of information and communication technology and social networking site use on older people's well-being in relation to loneliness: review of experimental studies. J Med Internet Res 2021 Mar 01;23(3):e23588 [FREE Full text] [doi: 10.2196/23588] [Medline: 33439127]
24. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester (UK): John Wiley & Sons; 2019.
25. Akers J. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York: CRD, University of York; 2009.
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
27. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
29. Egger M, Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
30. 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]

31. Johansson L, Hagman E, Danielsson P. A novel interactive mobile health support system for pediatric obesity treatment: a randomized controlled feasibility trial. BMC Pediatr 2020 Sep 23;20(1):447 [FREE Full text] [doi: 10.1186/s12887-020-02338-9] [Medline: 32967638]
32. Chai LK, Collins CE, May C, Ashman A, Holder C, Brown LJ, et al. Feasibility and efficacy of a web-based family telehealth nutrition intervention to improve child weight status and dietary intake: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul 31:1357633X19865855. [doi: 10.1177/1357633X19865855] [Medline: 31364474]
33. Nyström C, Sandin S, Henriksson P, Henriksson H, Maddison R, Löf M. A 12-month follow-up of a mobile-based (mHealth) obesity prevention intervention in pre-school children: the MINISTOP randomized controlled trial. BMC Public Health 2018 May 24;18(1):658 [FREE Full text] [doi: 10.1186/s12889-018-5569-4] [Medline: 29793467]
34. Bruñó A, Escobar P, Cebolla A, Álvarez-Pitti J, Guixeres J, Lurbe E, et al. Home-exercise childhood obesity intervention: a randomized clinical trial comparing print versus web-based (Move It) platforms. J Pediatr Nurs 2018 Sep;42:79-84. [doi: 10.1016/j.pedn.2018.04.008] [Medline: 29747957]
35. Rerksuppaphol L, Rerksuppaphol S. Internet based obesity prevention program for thai school children - a randomized control trial. J Clin Diagn Res 2017;11(3):7-11. [doi: 10.7860/jcdr/2017/21423.9368]
36. Mameli C, Brunetti D, Colombo V, Bedogni G, Schneider L, Penagini F, et al. Combined use of a wristband and a smartphone to reduce body weight in obese children: randomized controlled trial. Pediatr Obes 2018 Feb;13(2):81-87. [doi: 10.1111/ijpo.12201] [Medline: 27900849]
37. Nawi AM, Jamaludin FI. Effect of internet-based intervention on obesity among adolescents in Kuala Lumpur: a school-based cluster randomised trial. Malays J Med Sci 2015;22(4):47-56 [FREE Full text] [Medline: 26715908]
38. Abraham AA, Chow WC, So HK, Yip BH, Li AM, Kumta SM, et al. Lifestyle intervention using an internet-based curriculum with cell phone reminders for obese Chinese teens: a randomized controlled study. PLoS One 2015;10(5):e0125673 [FREE Full text] [doi: 10.1371/journal.pone.0125673] [Medline: 25946465]
39. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]
40. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):723-731 [FREE Full text] [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
41. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
42. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MF, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
43. Khokhar B, Jones J, Ronksley PE, Armstrong MJ, Caird J, Rabi D. Effectiveness of mobile electronic devices in weight loss among overweight and obese populations: a systematic review and meta-analysis. BMC Obes 2014;1:22 [FREE Full text] [doi: 10.1186/s40608-014-0022-4] [Medline: 26217509]
44. Kim HN, Seo K. Smartphone-based health program for improving physical activity and tackling obesity for young adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
45. 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;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
46. Park SH, Hwang J, Choi YK. Effect of mobile health on obese adults: a systematic review and meta-analysis. Healthc Inform Res 2019 Jan;25(1):12-26 [FREE Full text] [doi: 10.4258/hir.2019.25.1.12] [Medline: 30788177]
47. 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 1;181(5):337-348. [doi: 10.1093/aje/kwu260] [Medline: 25673817]
48. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462. [doi: 10.3945/an.116.014100] [Medline: 28507010]
49. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Woods HB. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
50. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
51. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]

52. Seo YG, Noh HM, Kim SY. Weight loss effects of circuit training interventions: a systematic review and meta-analysis. Obes Rev 2019 Nov 19;20(11):1642-1650. [doi: 10.1111/obr.12911] [Medline: 31322317]
53. Vizcaíno V, Aguilar F, Gutiérrez R, Martínez M, López M, Martínez S, et al. Assessment of an after-school physical activity program to prevent obesity among 9- to 10-year-old children: a cluster randomized trial. Int J Obes (Lond) 2008 Jan 25;32(1):12-22. [doi: 10.1038/sj.ijo.0803738] [Medline: 17895883]
54. Bugge A, El-Naaman B, Dencker M, Froberg K, Holme IM, McMurray RG, et al. Effects of a three-year intervention: the Copenhagen School Child Intervention Study. Med Sci Sports Exerc 2012 Jul;44(7):1310-1317. [doi: 10.1249/MSS.0b013e31824bd579] [Medline: 22297806]
55. Aguilar F, Martínez-Vizcaíno V, López M, Martínez M, Gutiérrez R, Martínez S, et al. Impact of an after-school physical activity program on obesity in children. J Pediatr 2010 Jul;157(1):36-42. [doi: 10.1016/j.jpeds.2009.12.046] [Medline: 20227726]
56. Martínez-Vizcaíno V, Sánchez-López M, Notario-Pacheco B, Salcedo-Aguilar F, Solera-Martínez M, Franquelo-Morales P, et al. Gender differences on effectiveness of a school-based physical activity intervention for reducing cardiometabolic risk: a cluster randomized trial. Int J Behav Nutr Phys Act 2014 Dec 10;11(1):154 [FREE Full text] [doi: 10.1186/s12966-014-0154-4] [Medline: 25491026]
57. 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]
58. Okorodudu DE, Bosworth HB, Corsino L. Innovative interventions to promote behavioral change in overweight or obese individuals: a review of the literature. Ann Med 2015 May;47(3):179-185 [FREE Full text] [doi: 10.3109/07853890.2014.931102] [Medline: 25011006]
59. Sjöberg RL, Nilsson KW, Leppert J. Obesity, shame, and depression in school-aged children: a population-based study. Pediatrics 2005 Sep 01;116(3):389-392. [doi: 10.1542/peds.2005-0170] [Medline: 16140683]
60. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]

Abbreviations

%BF: percentage body fat
BW: body weight
ICT: information and communication technology
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
WC: waist circumference
WMD: weighted mean difference

Edited by R Kukafka, G Eysenbach; submitted 22.03.21; peer-reviewed by J Piqueras, C Carrion; comments to author 26.04.21; revised version received 02.05.21; accepted 03.10.21; published 17.11.21.

Please cite as:
Park J, Park MJ, Seo YG
Effectiveness of Information and Communication Technology on Obesity in Childhood and Adolescence: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(11):e29003
URL: https://www.jmir.org/2021/11/e29003
doi:10.2196/29003
PMID:34787572

Original Paper

Clinical Effectiveness and Cost-effectiveness of Videoconference-Based Integrated Cognitive Behavioral Therapy for Chronic Pain: Randomized Controlled Trial

Kayoko Taguchi[1], PhD; Noriko Numata[1], PhD; Rieko Takanashi[2], PhD; Ryo Takemura[3], PhD; Tokiko Yoshida[1], PhD; Kana Kutsuzawa[4], MSc; Kensuke Yoshimura[5], MD, PhD; Natsuko Nozaki-Taguchi[5], MD, PhD; Seiji Ohtori[5], MD, PhD; Eiji Shimizu[4], PhD, MD

[1]Research Center for Child Mental Development, Chiba University, Chiba, Japan

[2]Department of Psychology, Teikyo University, Hachioji, Japan

[3]Keio University Hospital, Tokyo, Japan

[4]Graduate School of Medicine, Chiba University, Chiba, Japan

[5]Chiba University Hospital, Chiba, Japan

Corresponding Author:
Kayoko Taguchi, PhD
Research Center for Child Mental Development
Chiba University
Inohana 1-8-1
Chuo-ku
Chiba, 260-8670
Japan
Phone: 81 043 226 2027 ext 5208
Fax: 81 043 226 2028
Email: k.taguchi@chiba-u.jp

Abstract

Background: Cognitive behavioral therapy is known to improve the management of chronic pain. However, the components of this therapy are still being investigated and debated.

Objective: This study aimed to examine the effectiveness of an integrated cognitive behavioral therapy program with new components (attention-shift, memory work, video feedback, and image training) delivered via videoconferencing.

Methods: This study was unblinded and participants were recruited and assessed face-to-face in the outpatient department. We conducted a randomized controlled trial for chronic pain to compare 16 weekly videoconference-based cognitive behavioral therapy (vCBT) sessions provided by a therapist with treatment as usual (TAU). Thirty patients (age range, 22-75 years) with chronic pain were randomly assigned to either vCBT (n=15) or TAU (n=15). Patients were evaluated at week 1 (baseline), week 8 (midintervention), and week 16 (postintervention). The primary outcome was the change in pain intensity, which was recorded using the numerical rating scale at 16 weeks from the baseline. Secondary outcomes were pain severity and pain interference, which were assessed using the Brief Pain Inventory. Additionally, we evaluated disability, pain catastrophizing cognition, depression, anxiety, quality of life, and cost utility.

Results: In the eligibility assessment, 30 patients were eventually randomized and enrolled; finally, 15 patients in the vCBT and 14 patients in the TAU group were analyzed. Although no significant difference was found between the 2 groups in terms of changes in pain intensity by the numerical rating scale scores at week 16 from baseline (P=.36), there was a significant improvement in the comprehensive evaluation of pain by total score of Brief Pain Inventory (–1.43, 95% CI –2.49 to –0.37, df=24; P=.01). Further, significant improvement was seen in pain interference by using the Brief Pain Inventory (–9.42, 95% CI –14.47 to –4.36, df=25; P=.001) and in disability by using the Pain Disability Assessment Scale (–1.95, 95% CI –3.33 to –0.56, df=24; P=.008) compared with TAU. As for the Medical Economic Evaluation, the incremental cost-effectiveness ratio for 1 year was estimated at 2.9 million yen (about US $25,000) per quality-adjusted life year gained.

Conclusions: The findings of our study suggest that integrated cognitive behavioral therapy delivered by videoconferencing in regular medical care may reduce pain interference but not pain intensity. Further, this treatment method may be cost-effective, although this needs to be further verified using a larger sample size.

Trial Registration: University Hospital Medical Information Network UMIN000031124; https://tinyurl.com/2pr3xszb

(J Med Internet Res 2021;23(11):e30690) doi:10.2196/30690

KEYWORDS

cognitive behavioral therapy; chronic pain; medical economic evaluation; EQ-5D-5L; telemedicine

Introduction

Global reports indicate that chronic pain affects an estimated 20%-50% of people [1-4]. The COVID-19 pandemic has made it more difficult for people with chronic pain to safely visit hospitals. Chronic pain results in significant economic losses, with studies suggesting that pain management imposes a substantial burden on health care resources worldwide [5,6]. In Japan, it has been reported that economic loss from the inability to work due to pain is approximately 1.8 trillion yen per year [7]. It is, therefore, necessary to establish an urgent treatment system to benefit such patients and, in turn, the nation. Although multidisciplinary pain management is recommended as effective treatment for chronic pain, in clinical practice, psychosocial approaches to chronic pain are generally considered a last resort, thus delaying psychosocial intervention. Therefore, one can say that psychosocial approaches to chronic pain are underdeveloped or insufficient [8].

Cognitive behavioral therapy (CBT) is a structural intervention that encourages the transformation of a patient's cognition and behavior and directly addresses major psychological problems associated with chronic pain, such as repetitive ideation, concerns, emotions, and behaviors. CBT is recognized and recommended as an effective treatment for managing chronic pain [9]. The latest review on the effect of CBT verified by randomized controlled trials (RCTs) has added 41 new studies to the existing 34 studies, thus creating a large pool of 75 verified RCTs. Comparisons of CBT with active controls showed a slight benefit in terms of pain intensity, disability, and distress immediately after treatment [10]. It was also observed that there were small merits in each of the 3 outcomes as compared to no treatment. At follow-up, pain, disability, distress, and other variables were maintained in comparison with no treatment, but there is a lack of rigorous studies that involve active controls [10]. In summary, there is evidence of efficacy, although the effect size of CBT for chronic pain is small and insufficient.

The use of remote treatment across the internet has been increasing and its effectiveness has been demonstrated for various intractable diseases [11-14]. Internet-based CBT conducted for chronic pain has been on the rise [15-18]. While such remote treatment is known by several names such as internet-delivered CBT, web-based CBT, and telemedicine, these are all strictly different interventions. Web-based CBT and internet-delivered CBT were used synonymously in a study in which patients performed CBT on their own as self-help training or received regular therapist feedback that was not face-to-face [19,20]. Moreover, many studies involved treatment programs delivered via the internet or through smartphone apps, with the aim of self-management by patients (no intervention by the therapist) [21-23].

Only a few studies have examined the effectiveness of videoconference-based CBT (vCBT), which is face-to-face CBT using a videoconference system and not a general internet-delivered CBT for chronic pain management [24-26]. The aim of this study was to verify the cost and clinical effectiveness of a new integrated CBT program for chronic pain, delivered through videoconferencing (vCBT), and to compare with treatment as usual (TAU) for pain intensity, pain interference, disability, pain-related catastrophizing cognition, depression, anxiety, quality of life, and cost utility.

Methods

Study Design, Setting, and Participants

This study was designed as a prospective randomized unblinded pilot trial comparing vCBT as the intervention group to TAU as the control group at the academic outpatient clinic of the Cognitive Behavioral Therapy Center at Chiba University Hospital. Patients with intractable chronic pain, aged 18-75 years, were randomized and enrolled in one of the two groups. The intervention period was 16 weeks for both groups. The vCBT group received 1 weekly session of an integrated CBT program in addition to regular medical care. The TAU group continued outpatient consultation more than once for 8 weeks. For primary and secondary outcomes, patients were evaluated at week 1 (baseline), week 8 (midintervention), and week 16 (postintervention).

Although it is difficult to define intractable chronic pain, it is considered "drug-resistant" because almost all patients with acute pain receive pharmacotherapy. A review of evidence-based clinical trial designs for chronic pain pharmacotherapy states that "regulatory agencies such as the United States Food and Drug Administration and European Medicines Agency require studies of 12-weeks duration for chronic pain such as neuropathic pain to demonstrate the durability of response" [27]. Another review states that "some longer duration trials have shown efficacy of the investigational medication early in the course of treatment, only to lose statistical significance as the placebo group catches up" [28]. Consequently, the appropriate period of drug resistance cannot be determined, although taking these previous studies into account, the condition of drug resistance can be thought of as one in which individuals with chronic pain do not show moderate to remarkable improvement owing to poor tolerability, despite receiving sufficient pharmacotherapy for at least 12 weeks.

Ethics and Dissemination

This study was conducted with the approval of the Institutional Review Board of the Chiba University Hospital (approval ID G29049). In addition, the Clinical Research Ethics Review Committee oversaw the proper implementation of the test at least once a year. The trial registration number was University Hospital Medical Information Network UMIN000031124. The patients willing to participate in this study were informed of the study objectives and were asked for their consent to participate. Each patient was informed that participation was voluntary and full anonymity would be provided. Each patient was required to provide written consent for participation.

Recruitment

We recruited participants through web-based and newspaper advertisements from April 2018 to November 2019. Patient recruitment was announced by the doctors at outpatient clinics in the Department of Orthopedics and Pain Anesthesiology at Chiba University Hospital and in all medical institutions in the Chiba prefecture as well. All recruitment materials referred patients to our study website, which explained the study in detail. All participants who gave their permission to be enrolled in the study were required to continue treatment with their general practitioners as TAU. Patients who were interested in the study could inquire about the details via email. This mail was also used as an application form to ask patients to record their age, sex, condition of chronic pain, contact information, and so forth.

Eligibility Procedure for Participation and Diagnosis

Written informed consent was obtained face-to-face from all patients after they were fully briefed on the procedure. Following this, a screening eligibility assessment for inclusion and exclusion criteria was performed. Participants were asked to record their pain intensity rated by the Numerical Rating Scale (NRS) daily for a week. Inclusion criteria were as follows: (1) fulfillment of the criteria of somatic symptom disorder, with predominant pain according to the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), (2) age ranging from 18 to 75 years to avoid the risks from cognitive decline, (3) not showing moderate to remarkable improvement despite receiving sufficient pharmacotherapy for more than 8 weeks or due to poor tolerability, and (4) appropriate mental and physical conditions to maintain therapy. In the case of patients with depression or other anxiety conditions, they met the criteria to take part in the study if their pain was the primary impairment. The exclusion criteria were (1) comorbidity of serious mental disorders such as neurocognitive disorder, psychotic disorder, bipolar disorder, or substance-related disorder based on the criteria in DSM-5; (2) major pain caused by cancer; (3) if their pain did not interfere with their daily life (PDAS: Pain Disability Assessment Scale score of 9 or less); (4) mental retardation, neurocognitive disorders (dementia), and autism spectrum disorder; and (5) litigation or compensation concerning pain symptoms. In this study, patients were required to be able to use a videoconferencing system at home. In case of patients who did not have an internet connection in their houses, we rented tablet computers and mobile Wi-Fi devices for them.

Randomization and Case Registration

The eligible patients were randomly assigned to either the vCBT group or the TAU group using the minimization method used in clinical trials to ensure a balance in pain intensity score and gender. Per the allocation adjustment factor, the pain intensity score on NRS was allocated at 6.3 for each group [29]. The randomization and assignment were done by the Clinical Research Data Center, and the results were informed to the researchers by fax.

Intervention

The intervention period was from May 2018 to April 2020. In both groups, patients underwent 16 weeks of intervention and answered questions regarding primary and secondary outcomes at 1 week (baseline), 8 weeks (middle), and 16 weeks (post). All patients continued the regular medical care that they would normally receive as treatment. While participating in this study, no patient was permitted to seek any new treatment other than those that their primary care doctor ordered. In addition to regular medical care, those allocated to the vCBT group received weekly 50-minute sessions over 16 weeks of integrated CBT program using real-time internet videoconferencing. If the therapists or patients found it impossible to continue owing to adverse events, a gap of maximum 30 days during the intervention was permitted.

Videoconference-Based Integrated CBT Program

The CBT program that we adopted is an integrated CBT program that is longer than conventional interventions and consists of several new sessions not used in traditional CBT protocols. Our developed protocol with face-to-face CBT sessions provided by the videoconference system (web-based CBT) has been shown to improve catastrophic cognition, disability, and mood [30]. Conventional CBT programs for chronic pain often comprise 8-12 intervention sessions. In almost all programs, psychoeducation for pain, case formulation for understanding cognitive behavioral models of chronic pain, relaxation exercises such as breathing, and cognitive reconstruction, among others, were included [31]. Each of the 16 sessions lasted 50 minutes. We added 4 new sessions: tactile attention-shift training (session 4), memory work based on peak-end rule (session 10), sharpening behavioral image training (session 11), and video feedback (session 12) to the conventional CBT program (shown in Table 1).

Table 1. Integrated cognitive behavioral therapy program for chronic pain.

Session	Program	Description
Session 1	Introduction	Therapists explained the purpose of cognitive behavioral therapy and set short-, medium-, and long-term treatment goals.
Session 2	Psychoeducation	Patients studied ideas such as the mechanism of pain, gate-control theory, and acceptance of pain.
Session 3	Relaxation	Patients practiced progressive muscle relaxation and abdominal breathing techniques.
Session 4	Tactile attention-shift training	Patients practiced flexibly shifting their excessive attention to pain.
Session 5	Case formulation	Patients learned their own cognitive behavioral models and vicious pain-causing cycles.
Session 6	Safety behaviors	For behavioral activation, patients understood avoiding action due to pain and learned the demerits of continuing safety action such as avoidance, makeshift action.
Session 7	Cognitive restructuring 1	Patients' thinking habits were examined and they learned how to change their irrational thinking.
Session 8	Cognitive restructuring 2	Patients' thinking habits were examined and they learned how to change their irrational thinking.
Session 9	Activity pacing	Patients spaced out activities to manage pain.
Session 10	Memory work using the peak end rule	By reexamining their pain memory, patients learned that it influences chronic pain.
Session 11	Mental practice	Patients practiced imagining the movement of their body in pain and maintaining hope.
Session 12	Visual feedback	Patients performed mirror therapy as an alternative, recording own actions and observing ideal movement.
Session 13	Behavioral experiments 1	Patients practiced step by step those actions that could not be performed because of pain.
Session 14	Behavioral experiments 2	Patients practiced step by step those actions that could not be performed because of pain.
Session 15	Summary	We reviewed all the sessions and confirmed any remaining issues.
Session 16	Relapse prevention	Patients learned to think about how to respond when the pain recurred.

Tactile attention-shift training (session 4): Patients with chronic pain tend to demonstrate "attention bias." Excessive attention causes patients' pain to increase. Patients were trained to flexibly shift their attention. Furthermore, it has been found that pain may be relieved by gentle strokes that promote the secretion of oxytocin. This is referred to as the Science Touch method and suggests relief from chronic pain [32]. Attention-shift training takes place while being conscious of the sense of touch.

Memory work based on the peak-end rule (session 10): With significant chronic pain, patients tend to consider their memory of pain to be worse than the actual intensity of the pain. This is because the memory of an intense pain experience is saved as a traumatic memory [33]. The peak-end rule theory explains that "unpleasant experiences like pain are memorized as the average of the strength of the pain peak and the strength of the end" [33]. Patients with chronic pain tend to retain the memory of the peak of their pain experience, when the pain was too intense, and they may not be able to remember the end or release of their pain. Their peak (painful) memories are thus surpassed by the end (good) memories. Therefore, in their pain memory, the average degree of the pain and unpleasantness is high, and it remains as an extreme memory where objectivity about the pain condition is lost. This causes negative emotions such as anxiety and depressive feelings, which can be a factor for negative thoughts. This is considered catastrophic cognitionit is caused by pain but at the same time, it becomes the factor of chronicity. This session seeks to reduce pain by recalling memories of intense pain in detail and reconstructing them in objective memory.

Sharpening behavioral image training (session 11): We were inspired by sports training and poststroke rehabilitation [34,35] to compose this session. Many patients with chronic pain cannot imagine that they can move the injured body part since they have not moved it in a long time. Therefore, this session helped patients increase their daily movement using "image training" until they could clearly imagine moving their painful body parts themselves.

Video feedback (session 12): We developed this session in line with the principles of mirror therapy, a treatment used for phantom limb pain [36]. In this therapy, when the patient moves their healthy side, it is reflected in the mirror and the patient visually recognizes the movement. As a result, the brain interprets it as the amputated side that is moving and thus, the pain is alleviated [37,38]. When there is no movement in the painful area for a long time, it is considered to be a state in which the sensory and motor contact with the brain is severely disconnected. This is regarded as being similar to experiencing phantom limb pain. Thus, we considered that if pain can be alleviated by observing repeated normal movements in the

mirror, observing normal movements in videos should also have similar effects.

System Safety

In this study, we adopted the ISO 27001-certified Cisco WebEx as the internet conference system. Countermeasures against unauthorized access, information leakage, and others were taken, and safety problems were cleared.

Measures and Evaluation

Pretreatment Measures

After enrollment, we assessed the baseline characteristics of the patient's sex, age, education, marital status, comorbidity, employment status, age at onset of pain, duration of pain, and treatment history before they entered the intervention period (shown in Table 2). The following outcome measures were set based on the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials. Furthermore, "Assessment of Chronic Pain" was recommended by the International Pain Society [39].

Table 2. Baseline characteristics of the patients.

Characteristics	Videoconference-based cognitive behavioral therapy (n=15)	Treatment as usual (n=14)
Age (years), mean (SD)	50.0 (14.3)	43.9 (12.5)
Females, n (%)	10 (67)	9 (64)
Education history (years), mean (SD)	13.6 (2.6)	13.4 (3.8)
Currently employed, n (%)	4 (26)	5 (35)
Family members living together, mean (SD)	2.0 (2.1)	1.8 (1.0)
Chronic pain site, n (%)[a]		
Lower back	8 (53)	3 (21)
Back	6 (38)	3 (21)
Neck	4 (27)	1 (7)
Arm	3 (21)	4 (27)
Leg	7 (47)	7 (50)
Other	8 (53)	2 (14)
Duration of disease (years), mean (SD)	11.0 (12.6)	7.6 (5.8)
Mental status comorbidity, n (%)	6 (40)	8 (57)

[a]For chronic pain site, duplicate answers were possible.

Adverse Events

All adverse events were reported irrespective of their relevance to the intervention of this study, and serious adverse events were immediately reported to the institutional review board of the Chiba University Hospital and registered with the hospital risk management system as well. Moreover, an independent data monitoring committee accurately verified the detailed records of the clinical study's progress, critical efficacy variables, and safety data.

Primary Outcome Measures

The primary outcome was the change in the pain intensity from baseline to week 16, as indicated by the NRS score. The NRS is a self-rated questionnaire that measures pain intensity on a scale of 0-10, where 0=nothing and 10=severe. Patients were made to keep a daily pain diary. They recorded (1) maximum pain throughout the day, (2) minimum pain, (3) usual pain, and calculated the weekly average for pain on the day of the session (each NRS score=sum total of 1-week NRS score/7). Numerical values obtained by averaging the values in (1), (2), and (3) were taken as the main evaluation items comprising the composite value of NRS. The measurement has been shown to be reliable and valid [40].

Secondary Outcome Measures

All secondary outcomes were measured at 8 weeks and 16 weeks from the baseline.

Pain intensity: The secondary outcome was change in pain intensity (maximum, minimum, usual score) from baseline to week 8 on the NRS.

Comprehensive pain score: Comprehensive pain was assessed with the Japanese translation of the Brief Pain Inventory (BPI) [41]. BPI is composed of 2 factors: pain severity and pain interference. Pain severity means the pain intensity, but we use the original term as used in the BPI. The scale has a high reliability (coefficient alpha greater than .80) and established validity. Pain severity on the BPI comprises 4 items (worst, least, average, and current). They are assessed as 0 (nothing) to 10 (severe), with higher scores representing worse pain. Pain severity was calculated as the average of the 4 scores. Pain interference of BPI is a 7-item measure designed to assess pain interference by sleep, mood, social relations, and enjoyment of life. On an 11-point scale (0=does not interfere, 10=completely

interferes), patients indicated how much pain had interfered "in the past 24 hours" with different functional aspects. This score was the average of the 7 scores, and the total score was calculated as a composite score.

Cognition related to pain: Catastrophizing one's perception of pain was measured using the Pain Catastrophizing Scale (PCS). This scale has been shown to have high internal consistency (Cronbach α range .67-.87) [42]. The PCS comprises 13 items that evaluate the degree of catastrophizing cognition about pain. The responses are recorded on a 5-point Likert scale, where 0=not at all to 4=all the time. The total PCS scores ranged from 0 to 52, and the clinical cutoff value for the score was over 30 [42,43].

Disability: The degree of life disability due to pain was measured using PDAS. It is composed of 3 factors and supported by a high level of internal consistency (Cronbach α range .87-.95). It consists of 20 items on a 4-point Likert scale and is evaluated from 0 to 60 points, with a higher score indicating a higher degree of daily disability [44]. The clinical cutoff for PDAS is 10 points.

Depression and anxiety: Depressive symptoms were assessed with Beck's Depression Inventory II and Patient Health Questionnaire-9 items (PHQ-9). BDI-II has an internal consistency of approximately .90, and the test-retest reliability ranges from .73 to .96. It consists of 21 items with 4 response statements designed to assess the severity of current symptoms of depressive disorders. The total scores on the measure range from 0 to 63. Scores below 10 are regarded as reflecting "minimal or no" depression, whereas score ranges of 10-18, 19-29, and 30-63 reflect "mild to moderate," "moderate to severe," and "severe" depression, respectively [45,46]. PHQ-9 has diagnostic validity (for the diagnosis of any one or more PHQ disorders, κ=0.65; overall accuracy, 85%; sensitivity, 75%; specificity, 90%). It consists of 9 items scored on a 4-point Likert scale (0=not at all, 1=on several days, 2=half or more of days, and 3=almost daily). The minimum score is 0 and the maximum score is 27 (0-4, 5-9, 10-14, 15-19, and 20-27, indicating no, mild, moderate, moderate to severe, and severe symptoms, respectively) [47]. The PHQ cutoff score for clinically significant depressive symptoms is 10. Anxiety was measured on the Generalized Anxiety Disorder (GAD-7) scale, which has been shown to have reliability, criterion, construct, factorial, and procedural validity. Cutoff points that optimized sensitivity (89%) and specificity (82%) were identified. The scale has 7 items that assess the severity of GAD in the previous 2 weeks on a 4-point Likert scale (0=not at all, 1=one episode, 2=on half or more days, and 3=almost daily). The minimum score is 0 and the maximum score is 21 (0-4, 5-9, 10-14, and 15-21 indicating no, mild, moderate, and severe symptoms, respectively). The cutoff score for clinically significant symptoms of anxiety is 10 [48].

Health-related quality of life: European quality of life 5-dimensions 5-level (EQ-5D-5L) is a widely applied, valid, and reliable measure of quality of life. Its reliability was shown by Cronbach alpha (.70) [49]. EQ-5D-5L consists of 5 items related to mobility, self-care, common activities, pain/discomfort, and anxiety/depression. Patients answer each item on a scale of 1 to 5 (good to severe), and based on the score, the utility value, 0 to 1 (death to in good health), is calculated from the conversion device, which is used for medical economic evaluation [50,51].

Medical Economic Evaluation

All enrolled patients were asked to collect and submit receipts of their medical and drug expenses for chronic pain during this study (for 4 months). The incremental cost-effectiveness ratio (ICER) for the comparators vCBT versus TAU was estimated based on prepost evaluation (from baseline to 16 weeks), differences in medical costs, and effects of the condition. To estimate the cost utility of the intervention, quality-adjusted life years (QALYs) were calculated using the utility score of EQ-5D. To calculate the ICER, QALY values were calculated by multiplying one life year by the quality of life score. Quality of life scores were measured using the EQ-5D. We could not use quality of life scores 1 year (52 weeks) after the start of the intervention since the follow-up study had not been done. Therefore, we assumed that quality of life scores at week 16 could return to baseline at 1 year (week 52) as the minimum effect case. Conversely, we assumed that quality of life scores at week 16 could be the same as baseline at 1 year (week 52) as the maximum effect case.

Sample Size

In this study, we assumed that the difference in the amount of change in the NRS was 1.67 and the standard deviation was 1.8 and set the detection power to 80% and bilateral significance level to 5% in the two-sided *t* test. As a result, the required number of subjects per group was estimated to be almost 20. In the main analysis, analysis of covariance (ANCOVA) with the allocation factor as the covariate was used, and the detection power was calculated to be 82%. As this study is a pilot study, the number of cases was determined based on its feasibility.

Data Management

All data were properly managed by the submitting the case report form to the Clinical Research Data Center. In this center, researchers entered all the data using an access-log-restricted data system, which could be verified, and they created data sets. Independent data monitoring committees were regularly held and they performed risk-based monitoring. After the intervention was finished, the responsible doctors confirmed their data sets and locked the data. Then the locked data were transferred to the Pharmaceutical Statistics Office of the Department of Clinical Trials, Chiba University Hospital.

Statistical Analysis

Statistical analysis and reporting of this trial were conducted in accordance with the Consolidated Standards of Reporting Trials guidelines. Baseline variables were compared using Fisher exact test for categorical outcomes and an unpaired two-tailed *t* test for continuous variables. The significance level was set at .05. For the primary analysis of comparing treatment effects, the means of the least squares and their 95% CIs were estimated by ANCOVA with the change in the NRS composite score at week 16. This ANCOVA model took into account the variation caused by treatment effects, and gender and baseline NRS scores (≥6.3 and <6.3) were entered as covariates. As a sensitivity

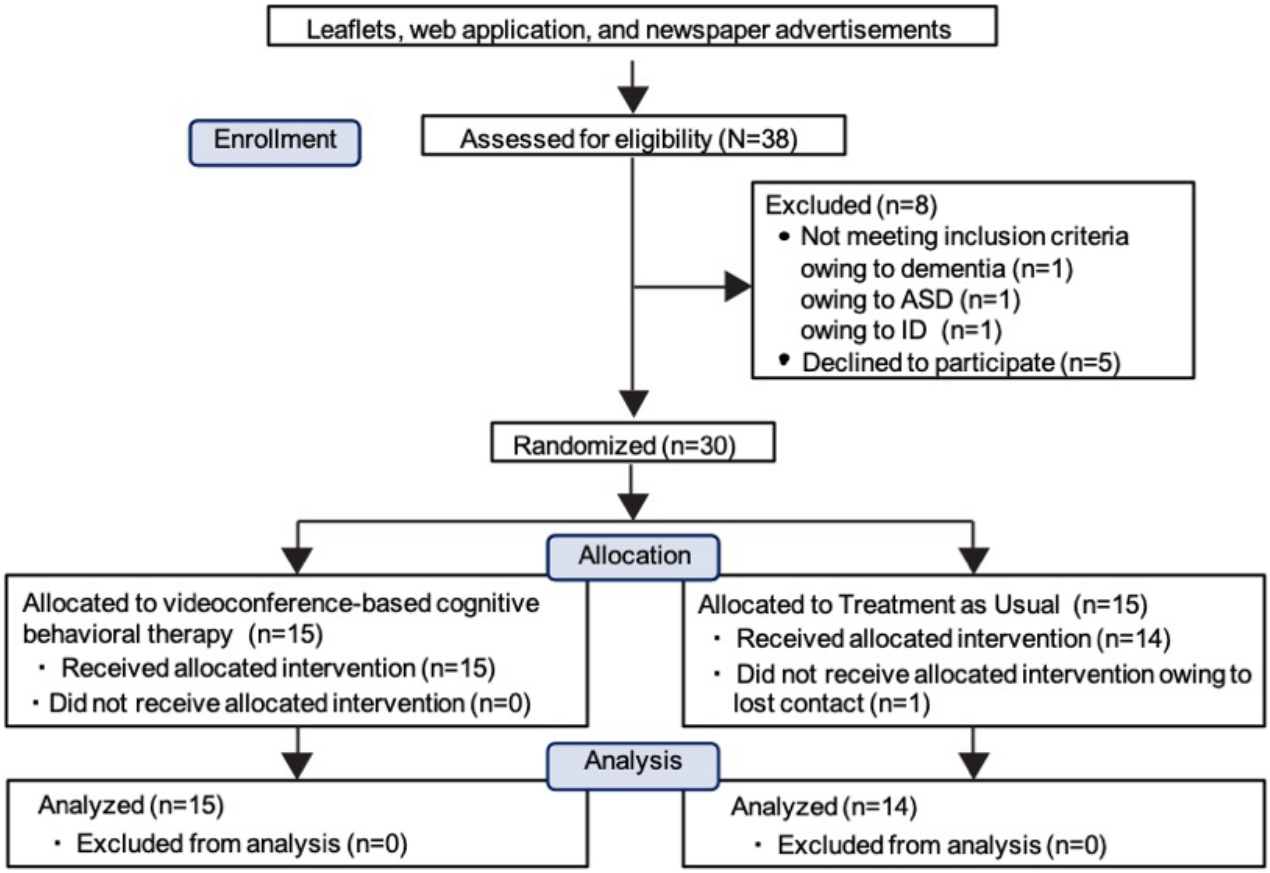
analysis, we showed the transition over time of the NRS scores of each group, confirmed the time course measurement data using the linear mixed effect model, and confirmed that it was not significantly different from the covariance analysis result. All comparisons were planned, and all *P* values were two-sided. *P* values <.05 were considered statistically significant. No special complementary processing by statistical methods was performed for the missing values. However, if necessary, complementary analysis using Mixed Model Repeated Measure was carried out for exploratory analysis. This analysis plan was created before the trial was started, and the protocol was approved. All statistical analyses were performed using SAS V.9.4. (SAS Institute, Cary).

Results

Participants

Figure 1 depicts the participant enrollment for this study. The recruitment process resulted in a total of 38 participants. Three participants were excluded owing to dementia (n=1), autism spectrum disorder (n=1), suspected intellectual disability (n=1), and 5 declined to participate. In the eligibility assessment, 30 patients were eventually enrolled and randomized.

Figure 1. Flowchart of participant selection in the trial. ASD: autism spectrum disorder; ID: intellectual disability.



Patients' Demographic and Clinical Characteristics

A total of 93% of the patients (14/15) completed all 16 sessions of the vCBT program and participated in the intervention throughout its duration; 2 patients missed 1 session each due to adverse events (see below). No patients in the vCBT group were excluded from the analysis. One patient in the TAU group did not report receiving regular medical treatment at week 8 and was excluded from the analysis. Finally, 15 patients in the vCBT and 14 patients in the TAU group were analyzed.

Adverse Events

In this study, 4 adverse events were reported by 4 different patients in the vCBT group. The first patient was hospitalized owing to worsening Behcet disease and declined to participate in the study for the fourth session. The second patient had sudden difficulty in opening his eyes due to medically unexplained eyelid pain and declined to participate in the study for the fifth session. The third patient had common cold. The fourth patient had temporomandibular joint disorders.

Table 2 shows the demographic data of the patients. There were no significant differences between vCBT and TAU in age, gender, length of education, employment status, and number of families living together (*P*=.24, *P*>.99, *P*=.73, *P*=.70, *P*=.73; respectively). Duration of illness in vCBT (mean 11.03 [SD 12.64] years) was significantly longer than that in TAU (mean 7.56 [SD 5.84] years). In both groups, more than 65% (19/29) of the patients were women. All patients in the vCBT group had an education period of 12 years or more (high school graduation or above), and TAU group had also 12 years or more except for one patient (9 years). There was no significant difference between the two groups. At baseline, nearly 70% (20/29) of patients were not working (vCBT 11/15, 55% vs TAU 9/14, 45%). Patients were living with at least one family, and only 1 patient receiving vCBT lived alone. The most

reported site of chronic pain was lower back pain. Many patients had orthopedic pain, while others had oral pain such as tongue pain, toothache, and general pain such as rheumatism and fibromyalgia.

Primary Outcomes

Table 3 shows the adjusted mean reductions of NRS in vCBT and TAU at 16 weeks (primary outcome) and at 8 weeks (secondary outcome) from the baseline. No significant difference was found between the 2 groups in terms of changes in composite NRS scores at week 16 from baseline (P=.36). Table 4 shows the raw data on the means and standard deviations of NRS scores in vCBT and TAU at 16 weeks and 8 weeks.

Table 3. Results of the efficacy on pain intensity in videoconference-based cognitive behavioral therapy versus treatment as usual.

Variable	At 8 weeks				At 16 weeks			
	Estimated amount of change	SE	*P* value	95% CI	Estimated amount of change	SE	*P* value	95% CI
Numerical rating scale (composite)	–0.28	0.49	.57	–1.29 to 0.72	–0.46	0.49	.36	–1.47 to 0.55
Numerical rating scale (max)	–0.13	0.54	.82	–1.24 to 0.99	–0.08	0.55	.88	–1.21 to 1.05
Numerical rating scale (min)	–0.29	0.54	.60	–1.41 to 0.83	–1.04	0.55	.07	–2.17 to 0.09
Numerical rating scale (usual)	–0.43	0.60	.47	–1.66 to 0.79	–0.04	0.60	.94	–1.27 to 1.18

Table 4. Change in numerical rating scale scores at the first, eighth, and sixteenth week.

Variable	At 1 week				At 8 weeks				At 16 weeks			
	vCBT[a]		TAU[b]		vCBT		TAU		vCBT		TAU	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
NRS[c] (composite)	15	5.67 (1.62)	14	5.01 (1.37)	13	5.42 (1.99)	14	4.92 (1.27)	13	5.08 (2.33)	14	4.76 (1.78)
NRS (max)	15	7.30 (1.42)	14	7.40 (1.40)	13	7.08 (1.49)	14	7.18 (1.77)	13	6.42 (2.18	13	6.38 (2.11)
NRS (min)	15	4.33 (2.04)	14	3.01 (1.71)	13	4.22 (2.19)	14	3.20 (1.44)	13	3.74 (2.43)	13	3.52 (1.95)
NRS (usual)	15	5.40 (1.77)	14	4.45 (1.45)	13	5.00 (2.58)	14	4.41 (1.07)	13	5.37 (2.20)	14	4.39 (1.85)

[a]vCBT: videoconference-based cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]NRS: numerical rating scale.

Secondary Outcomes

Tables 5 and 6 show the results of efficacy on the secondary outcomes. *P* values lesser than .05 were considered statistically significant. No significant difference was found between the 2 groups regarding changes in maximum, minimum, and usual NRS scores at week 16 from the baseline. In addition, there was no significant difference in the changes in all NRS scores at week 8 from baseline.

Comprehensive pain score: The adjusted mean reduction of the total BPI score in vCBT was significantly larger than that of TAU at 16 weeks (–1.43, 95% CI –2.49 to –0.37, *df*=24; *P*=.01). There was no significant difference between the 2 groups in the adjusted mean reduction of BPI pain severity at 16 weeks. The adjusted mean reduction of BPI pain interference in vCBT was significantly larger than that of TAU at 16 weeks (–1.95, 95% CI –3.33 to –0.56, *df*=24; *P*=.008). There was no significant difference between the 2 groups in the adjusted mean reduction of BPI total score, pain severity, and pain interference at 8 weeks.

Disability: The adjusted mean reduction of the PDAS score in vCBT was significantly larger than that of TAU at 16 weeks (–9.42, 95% CI –14.47 to –4.36, *df*=25; *P*=.001). There was no significant difference between the 2 groups in the adjusted mean reduction of PDAS at 8 weeks.

Catastrophizing cognition: The adjusted mean reduction of the PCS score in vCBT was larger than that of TAU at 16 weeks, although the difference was not statistically significant (–6.25, 95% CI –12.89 to 0.38, *df*=25; *P*=.07). There was no significant difference between the 2 groups in the adjusted mean reduction of PCS at 8 weeks.

Depression and anxiety: There was no significant difference between the 2 groups in the adjusted mean reduction of BDI-II, PHQ-9, and GAD-7 at 16 weeks and 8 weeks, respectively.

Health-related quality of life: The adjusted mean change of the EQ-5D-5L index score in vCBT was larger than that of TAU at 16 weeks, although the difference was not statistically significant (0.09, 95% CI –12.89 to 0.38; *P*=.06). There was also no significant difference between the 2 groups in the adjusted mean change of the EQ-5D-5L index score at 8 weeks.

Table 5. Results of the efficacy on secondary outcomes in videoconference-based cognitive behavioral therapy versus treatment as usual.

Outcome	At 8 weeks				At 16 weeks			
	Estimate	SE	*P* value	95% CI	Estimate	SE	*P* value	95% CI
BPI-J[a] (total)	–0.52	0.51	.33	–1.57 to 0.54	–1.43	0.51	.01	–2.49 to –0.37
BPI-J (pain severity)	0.19	0.46	.68	–0.76 to 1.14	–0.38	0.46	.42	–1.33 to 0.57
Pain worst	–0.25	0.56	.66	–1.41 to 0.91	–1.19	0.56	.05	–2.35 to –0.03
Pain least	1.15	0.55	.05	0.02 to 2.27	0.32	0.55	.56	–0.80 to 1.45
Pain average	0.21	0.55	.70	–0.92 to 1.35	–0.41	0.55	.47	–1.55 to 0.73
Pain current	–0.20	0.69	.77	–1.61 to 1.21	–0.11	0.69	.87	–1.52 to 1.30
BPI-J (pain interference)	–0.88	0.67	.20	–2.26 to 0.51	–1.95	0.67	.01	–3.33 to –0.56
PDAS[b]	–2.71	2.45	.28	–7.77 to 2.34	–9.42	2.45	<.001	–14.47 to –4.36
PCS[c]	–4.20	3.22	.20	–10.84 to 2.44	–6.25	3.22	.06	–12.89 to 0.38
BDI-II[d]	2.33	2.63	.38	–3.08 to 7.74	–2.01	2.63	.45	–7.42 to 3.39
PHQ-9[e]	–0.59	1.47	.69	–3.62 to 2.44	–2.15	1.47	.16	–5.18 to 0.89
GAD-7[f]	0.87	1.4	.54	–2.02 to 3.75	–1.28	1.4	.37	–4.16 to 1.61
EQ-5D-5L[g]	–0.03	0.04	.57	–0.11 to 0.06	0.09	0.04	.05	0.00 to 0.18

[a]BPI-J: Brief Pain Inventory-Japanese translation.

[b]PDAS: Pain Disability Assessment Scale.

[c]PCS: Pain Catastrophizing Scale.

[d]BDI-II: Beck's Depression Inventory II.

[e]PHQ-9: Patient Health Questionnaire-9 items.

[f]GAD-7: Generalized Anxiety Disorder Scale-7 items.

[g]EQ-5D-5L: European quality of life 5-dimensions 5-level.

Table 6. Change in the secondary outcomes in the first, eighth, and sixteenth week.

Variable	Preintervention (1st week)				Midintervention (8th week)				Postintervention (16th week)			
	vCBT[a]		TAU[b]		vCBT		TAU		vCBT		TAU	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
BPI-J[c] (total)	14	5.42 (1.74)	14	5.32 (1.27)	13	4.89 (1.71)	14	5.24 (1.33)	13	3.98 (1.83)	14	5.30 (1.63)
BPI-J (pain severity)	15	5.75 (1.35)	14	5.43 (1.70)	13	5.71 (1.71)	14	5.27 (1.02)	13	4.92 (1.84)	14	5.05 (1.36)
Pain worst	15	8.07 (1.22)	14	8.14 (1.56)	13	7.54 (1.05)	14	7.79 (1.42)	13	6.38 (1.80)	14	7.57 (1.83)
Pain least	15	3.60 (1.64)	14	2.71 (2.61)	13	4.15 (2.34)	14	2.43 (1.34)	13	3.62 (2.02)	14	2.71 (1.49)
Pain average	15	5.67 (1.45)	14	5.57 (1.83)	13	5.38 (1.94)	14	5.07 (1.21)	13	4.69 (1.97)	14	5.00 (1.62)
Pain current	15	5.67 (2.02)	14	5.29 (2.16)	13	5.77 (2.31)	14	5.79 (1.67)	13	5.00 (2.27)	14	4.93 (2.02)
BPI-J (pain interference)	14	5.18 (2.35)	14	5.27 (1.47)	13	4.42 (2.19)	14	5.22 (1.83)	13	3.44 (2.05)	14	5.44 (1.97)
PDAS[d]	15	30.47 (10.35)	14	27.14 (9.29)	13	24.00 (8.55)	14	23.93 (8.65)	13	19.15 (10.04)	14	25.79 (12.24)
PCS[e]	15	28.87 (10.39)	14	34.50 (6.26)	13	24.15 (10.27)	14	32.71 (8.00)	13	21.38 (10.59)	14	32.00 (10.78)
PHQ-9[f]	15	11.13 (5.64)	14	11.29 (6.22)	13	9.77 (4.95)	14	10.36 (5.17)	13	8.00 (5.15)	14	10.14 (5.80)
GAD-7[g]	15	6.13 (4.96)	14	8.71 (4.78)	13	6.69 (4.35)	14	7.21 (4.25)	13	4.69 (3.64)	14	7.36 (5.17)
BDI-II[h]	15	18.67 (7.66)	14	19.64 (8.87)	13	17.38 (10.74)	14	16.64 (10.63)	13	13.54 (8.32)	14	17.14 (10.31)
EQ-5D-5L[i]	15	0.54 (0.17)	14	0.55 (0.11)	13	0.57 (0.24)	14	0.61 (0.09)	13	0.60 (0.22)	14	0.52 (0.17)

[a]vCBT: videoconference-based cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]BPI-J: Brief Pain Inventory-Japanese translation.

[d]PDAS: Pain Disability Assessment Scale.

[e]PCS: Pain Catastrophizing Scale.

[f]PHQ-9: Patient Health Questionnaire-9 items.

[g]GAD-7: Generalized Anxiety Disorder Scale-7 items.

[h]BDI-II: Beck's Depression Inventory II.

[i]EQ-5D-5L: European quality of life 5-dimensions 5-level.

Medical Economic Evaluation

Complete data for the direct medical economic evaluation were available for 12 (80%) patients in the vCBT group and 12 (80%) patients in the TAU group with their informed consent. The data consisted of medical expenses covered by public health care insurance, prescription drug prices, and uncovered health care expenses for chronic pain during the intervention period (16 weeks or 4 months). The mean total cost per person for 4 months was 145,846 yen (1 USD=114.05 yen) in vCBT and 169,312 yen in the TAU group, thereby showing no significant difference between the 2 groups (P=.73) (Table 7). To calculate the ICER, we extracted incremental effects from EQ-5D and incremental costs from collected receipts of medical and drug expenses. The incremental effect of vCBT on TAU was determined to be 0.033 to 0.064 QALY for 1 year. We adopted 0.033 QALY as the incremental effect as the best case. Regarding the incremental cost, the total cost of 16 sessions of the vCBT program was 96,000 yen per person with 6000 yen per session. Therefore, we adopted 96,000 yen as the incremental cost for vCBT compared with TAU. A cost-benefit analysis based on the data showed that the ICER was almost 2.9 million (yen/QALY=96,000 yen divided by 0.033 QALY). Previous research that used willingness to pay (WTP) to obtain the criterion of the ICER showed that 5.0 million yen (US $48,158) per QALY gained is considered an acceptable threshold in Japan [52]. Therefore, it was suggested that vCBT is more cost-effective than TAU, even though it was a minimum effect case, as the improvement in quality of life returned to the baseline in 1 year (shown in Figure 2).

Table 7. Comparison of costs (in yen) for treatment of chronic pain.[a]

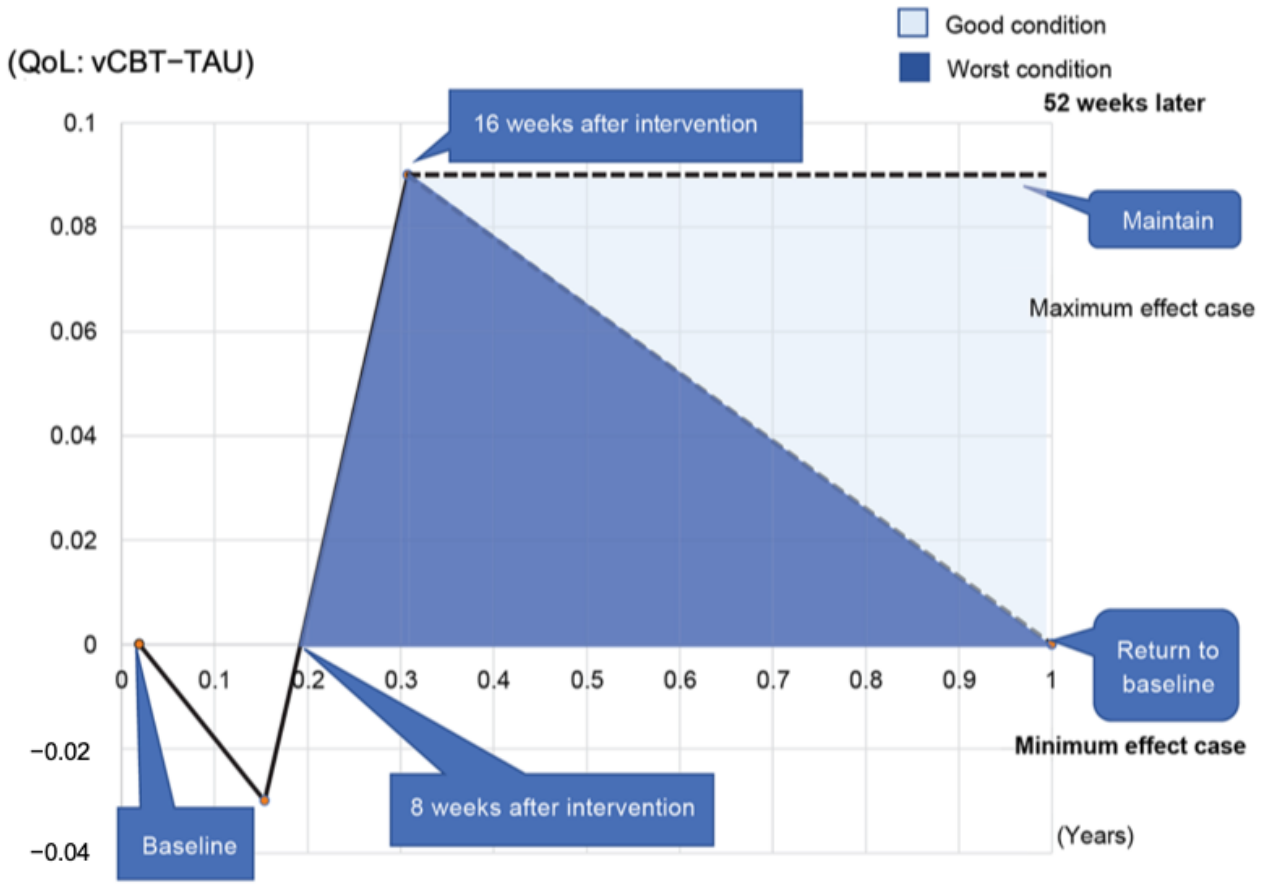
Costs	1st month	2nd month	3rd month	4th month	Total amount	*P* value
Insurance medical treatment for chronic pain (100% burden), mean (SD)						.69
TAU[b] (n=12)	16,057 (16,232)	15,898 (16,711)	10,681 (14,060)	18,014 (23,755)	60,650 (64,292)	
vCBT[c] (n=12)	9309 (8754)	19,182 (24,923)	28,728 (42,755)	17,153 (33,133)	74,372 (98,889)	
Outpatient prescription amount for chronic pain, mean (SD)						.99
TAU (n=12)	13,503 (14,082)	16,373 (17,924)	12,163 (11,756)	19,517 (23,939)	61,556 (58,758)	
vCBT (n=12)	23,129 (28,480)	8149 (10,933)	15,578 (15,318)	15,078 (14,047)	61,933 (59,961)	
Private insurance amount for chronic pain, mean (SD)						.35
TAU (n=12)	9637 (23,676)	12,049 (29,535)	11,710 (32,480)	13,710 (41,539)	47,106 (132,486)	
vCBT (n=12)	1817 (4051)	2233 (6249)	3811 (8236)	1680 (3723)	9541 (21,859)	
Total cost for chronic pain, mean (SD)						.73
TAU (n=12)	39,197 (39,425)	44,319 (46,594)	34,554 (50,316)	51,241 (63,891)	169,312 (195,072)	
vCBT (n=12)	34,255 (29,983)	29,564 (23,755)	48,116 (45,076)	33,910 (35,109)	145,846 (123,008)	

[a]1 USD=114.05 yen.

[b]TAU: treatment as usual.

[c]vCBT: videoconference-based cognitive behavioral therapy.

Figure 2. Difference in quality-adjusted life year between both groups at 1 week, 8 weeks, and 16 weeks, and prediction of difference in quality-adjusted life year after 1 year. QALY: quality-adjusted life year; TAU: treatment as usual; vCBT: videoconference-based cognitive behavioral therapy.



Discussion

To our knowledge, this study is the first RCT to evaluate the effectiveness of vCBT for chronic pain. In addition, we performed a cost-effectiveness analysis of the CBT program by using ICER. Although composite pain intensity by the NRS did not change, vCBT significantly improved the total BPI score, especially pain interference and disability in daily life. Furthermore, in the medical economic evaluation, although our

results showed no statistically significant difference, it has been suggested that vCBT may be more cost-effective than TAU for various reasons that we discuss in the following sections.

The latest systematic review of 75 RCTs about the effectiveness of face-to-face CBT has shown that CBT versus TAU at the treatment end for pain, disability, and distress showed a small effect size, and versus active treatment showed few effects (n=9401) [9]. In general, internet-based CBT is delivered through a computer application for administering self-management programs without a therapist. According to a systematic review of internet-based CBT for depression, anxiety disorder, and other functional disorders, including chronic pain, the CBT for chronic pain showed small to medium effect size from 10 RCT studies with low levels of evidence. Conversely, internet-based CBT for depression, social anxiety disorder, and panic disorder was classified as established, that is, they met the highest level of evidence criteria [11]. A systematic review of internet-based CBT for lower back pain reported high effectiveness for pain intensity and disability in 9 RCT studies (n=1796) [53]. Although there are several studies on internet-based CBT, there are only a few RCTs on vCBT, and almost all of them focus on self-care by patients. Only a pilot study of hematopoietic stem cell transplantation has reported on the feasibility and improved self-efficacy of vCBT [26].

As reported by multiple internet-based CBT studies, use of self-care programs or apps, treatments in which specialists check progress every other week, or telephone sessions have been demonstrated to be effective. Such internet-based CBT does not require multiple therapists, and interventions can be conducted for many patients concurrently. However, there is insufficient evidence for the effectiveness of the treatment. Chronic pain is unique to each individual and therefore, there is a need to consider stepped care for chronic pain, which is a system of delivering an evidence-based staged approach and provides treatment that matches individual needs and status. Ideally, it is desirable for patients to receive internet-based CBT in the early stage as first aid, such as through a self-care smartphone app. Moreover, patients who do not respond well to first aid should receive vCBT as the second aid.

vCBT is intended for the treatment of higher levels of chronic pain. We need to examine treatments for several levels of pain for treating complex and intractable chronic pain. This was the first RCT on vCBT showing significant improvement in interference (distress) and disability. Our primary outcome, pain intensity as measured using NRS, was not significantly different; however, we adopted the BPI total score as a comprehensive pain score, which showed significant results. The reason for the NRS result could be that TAU also improved pain intensity. According to the BPI total score, pain interference appeared to greatly improve, while pain severity only showed a slight change, which comprehensively improved the total score. We hypothesized that patients who received CBT increased their previously avoided daily movement by doing the CBT-related homework and confronting their pain, thus maintaining their pain intensity. Conversely, it was possible for patients with TAU to avoid several behaviors owing to pain; as a result, their pain interference score worsened gradually. Considering this development, a composite score such as a BPI total score should be adopted for assessing the effectiveness of treatment for chronic pain. A recent consensus recommended the use of comprehensive indicators such as a BPI [54]. In particular, the composite score of BPI is the right choice because it comprises 2 domains, pain severity and interference due to pain, which are measured as outcomes in all chronic pain clinical trials. Few studies have set the BPI total score as primary [23] and there might be a need to set a comprehensive scale, instead of a single rating scale, as the primary outcome in future research.

We expected improvement of cognition-related pain-related catastrophic thoughts, anxiety, and depressive symptoms, because a previous review had shown small-to-moderate effect size of catastrophic thinking and mood [9]. Moreover, one of the new additional sessions of our protocol focused on emotions related to pain (session 10). Contrary to expectations, there was no significant difference between PCS and mood. However, if the sample size is a little large, it may show a positive result as the PCS score of vCBT improves compared to TAU (shown in Table 6). Considering the previous review, it is not easy to improve mood by CBT for chronic pain [55]. Our new protocol also did not treat their anxiety and depression.

We thus added a new session to address emotions such as anxiety and depression. However, in this session, many patients could not recognize the emotions related to pain. They thought of their pain and emotions as separate and did not acknowledge emotions that accompanied the pain memory or emotions that accompanied the moment pain was felt. An RCT found that emotion-focused exposure intervention (10-15 sessions) compared to normal CBT for chronic pain is effective for catastrophic cognition and depression, in addition to other outcomes [56]. The research concluded that the emotion-focused approach is an effective option. We had a single emotion-focused session, which was too little time for patients to learn how to treat their emotions.

In medical economic evaluation, a cost-benefit analysis showed that ICER was almost 2.9 million (yen/QALY), which was lesser than Japan's WTP of 5 million yen. Furthermore, it was also shown to be lesser than the British WTP of £23,000 (approximately 3 million yen) and the United States WTP of $62,000 (approximately 6.7 million yen) [51]. At postintervention (16 weeks), the total cost of each group was not significantly different; however, the cost of vCBT did not change even though TAU increased gradually. In particular, the decrease in outpatient prescriptions of vCBT increased TAU costs, which may have affected the overall costs. We understand that the results of economic assessments are somewhat uncertain for a variety of reasons. Since this was a pilot study, the number of verifications was small. In addition, since no follow-up data were taken, the cost-effectiveness results were derived from assumptions of the sustainability of therapeutic efficacy. This uncertainty should be kept in mind if clinicians and policymakers set treatment guidelines or offer this program in a clinical setting.

This study has several limitations. First, the sample size was relatively small. In addition, this study was performed as a single-center study at our hospital. In the near future, large-scale multicenter trials are necessary. Second, further studies targeting

patients with specific types of chronic pain will be required to examine the effectiveness of this vCBT program. Third, because we did not use a psychological placebo group as a control condition, we were unable to control nonspecific factors and unravel the concrete effects of the vCBT program. Finally, the lack of follow-up data limited the generalizability of the conclusions. Long-term follow-up studies should be conducted in the future.

Despite the problems in this study, vCBT has the possibility of usability and effectiveness for patients with chronic pain. During COVID-19, in particular, vCBT may provide safety and alleviate the patient's stress through face-to-face counseling. In the future, we may study other pandemics and help patients to decrease their anxiety and symptoms by using internet-based CBT.

In summary, videoconference-based integrated CBT for chronic pain could improve pain interference and disability in daily life, and as a result, could even relieve comprehensive pain. Therefore, this program can be a valuable addition to regular medical care. Further research is needed to examine augmentation strategies, including an examination of the components of CBT that are best suited for different types of pain. Furthermore, while this study indicated the cost-effectiveness of this treatment within a small sample, this needs to be verified with a larger sample size.

Acknowledgments

We thank the Department of Clinical Trials and the Medical Research Data Center of Chiba University Hospital. We thank our research nurses, Hiroe Ota and Yoko Suwabe, for their huge support and our therapists, Yoko Okamoto and Yoichi Seki, who provided cognitive behavioral therapy in this study. We would also like to thank Editage (www.editage.com) for English language editing. Funding for this study was provided by the Health and Labor Sciences Research Grant (H29, Refractory, General, 062). The funding body had no role in the design of the study, collection, analysis, and interpretation of data and manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 134 KB - jmir_v23i11e30690_app1.pdf]

References

1. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
2. de Souza JB, Grossmann E, Perissinotti DMN, de Oliveira Junior JO, da Fonseca PRB, Posso IDP. Prevalence of Chronic Pain, Treatments, Perception, and Interference on Life Activities: Brazilian Population-Based Survey. Pain Res Manag 2017;2017:4643830 [FREE Full text] [doi: 10.1155/2017/4643830] [Medline: 29081680]
3. Goldberg DS, McGee SJ. Pain as a global public health priority. BMC Public Health 2011 Oct 06;11:770 [FREE Full text] [doi: 10.1186/1471-2458-11-770] [Medline: 21978149]
4. Treede R, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain 2015 Jun;156(6):1003-1007 [FREE Full text] [doi: 10.1097/j.pain.0000000000000160] [Medline: 25844555]
5. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8(1):8-20. [doi: 10.1016/j.spinee.2007.10.005] [Medline: 18164449]
6. Phillips CJ. The Cost and Burden of Chronic Pain. Rev Pain 2009 Jun;3(1):2-5 [FREE Full text] [doi: 10.1177/204946370900300102] [Medline: 26526940]
7. Takura T, Ushida T, Kanchiku T, Ebata N, Fujii K, DiBonaventura MD, et al. The societal burden of chronic pain in Japan: an internet survey. J Orthop Sci 2015 Jul;20(4):750-760. [doi: 10.1007/s00776-015-0730-8] [Medline: 25963609]
8. Hylands-White N, Duarte RV, Raphael JH. An overview of treatment approaches for chronic pain management. Rheumatol Int 2017 Jan;37(1):29-42. [doi: 10.1007/s00296-016-3481-8] [Medline: 27107994]
9. Williams ACDC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11:CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
10. Williams ACDC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020 Aug 12;8:CD007407. [doi: 10.1002/14651858.CD007407.pub4] [Medline: 32794606]
11. de Zwaan M, Herpertz S, Zipfel S, Svaldi J, Friederich H, Schmidt F, et al. Effect of Internet-Based Guided Self-help vs Individual Face-to-Face Treatment on Full or Subsyndromal Binge Eating Disorder in Overweight or Obese Patients: The INTERBED Randomized Clinical Trial. JAMA Psychiatry 2017 Oct 01;74(10):987-995 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2150] [Medline: 28768334]

12. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up: A Randomized Clinical Trial. JAMA Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
13. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, García Campayo J, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: Results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
14. Spence SH, Donovan CL, March S, Kenardy JA, Hearn CS. Generic versus disorder specific cognitive behavior therapy for social anxiety disorder in youth: A randomized controlled trial using internet delivery. Behav Res Ther 2017 Mar;90:41-57. [doi: 10.1016/j.brat.2016.12.003] [Medline: 27988427]
15. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
16. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
17. Friedberg F, Williams DA, Collinge W. Lifestyle-oriented non-pharmacological treatments for fibromyalgia: a clinical overview and applications with home-based technologies. J Pain Res 2012;5:425-435 [FREE Full text] [doi: 10.2147/JPR.S35199] [Medline: 23166446]
18. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
19. Janse A, Worm-Smeitink M, Bleijenberg G, Donders R, Knoop H. Efficacy of web-based cognitive-behavioural therapy for chronic fatigue syndrome: randomised controlled trial. Br J Psychiatry 2018 Feb;212(2):112-118. [doi: 10.1192/bjp.2017.22] [Medline: 29436329]
20. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-Based Cognitive Behavior Therapy for Depression in People With Diabetes Mellitus: A Randomized Controlled Trial. J Med Internet Res 2017 May 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
21. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
22. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
23. Kroenke K, Krebs EE, Wu J, Yu Z, Chumbler NR, Bair MJ. Telecare collaborative management of chronic pain in primary care: a randomized clinical trial. JAMA 2014 Jul 16;312(3):240-248. [doi: 10.1001/jama.2014.7689] [Medline: 25027139]
24. Colleen L, Beselli C. 4. Impacting healthcare outcomes: Connecting digital health within the veteran's health administration Home-Based Primary Care (HBPC) program. In: Three Facets of Public Health and Paths to Improvements. CA, USA: Academic Press; 2020:95-126.
25. Palyo SA, Schopmeyer KA, McQuaid JR. Tele-pain management: use of videoconferencing technology in the delivery of an integrated cognitive-behavioral and physical therapy group intervention. Psychol Serv 2012 May;9(2):200-202. [doi: 10.1037/a0025987] [Medline: 22662734]
26. Somers TJ, Kelleher SA, Dorfman CS, Shelby RA, Fisher HM, Rowe Nichols K, et al. An mHealth Pain Coping Skills Training Intervention for Hematopoietic Stem Cell Transplantation Patients: Development and Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Mar 19;6(3):e66 [FREE Full text] [doi: 10.2196/mhealth.8565] [Medline: 29555620]
27. Quessy SN, Rowbotham MC. Placebo response in neuropathic pain trials. Pain 2008 Sep 15;138(3):479-483. [doi: 10.1016/j.pain.2008.06.024] [Medline: 18706762]
28. Dworkin RH, Turk DC, Katz NP, Rowbotham MC, Peirce-Sandner S, Cerny I, et al. Evidence-based clinical trial design for chronic pain pharmacotherapy: a blueprint for ACTION. Pain 2011 Mar;152(3 Suppl):S107-S115. [doi: 10.1016/j.pain.2010.11.008] [Medline: 21145657]
29. Ang DC, Jensen MP, Steiner JL, Hilligoss J, Gracely RH, Saha C. Combining cognitive-behavioral therapy and milnacipran for fibromyalgia: a feasibility randomized-controlled trial. Clin J Pain 2013 Sep;29(9):747-754 [FREE Full text] [doi: 10.1097/AJP.0b013e31827a784e] [Medline: 23446065]
30. Taguchi K, Numata N, Takanashi R, Takemura R, Yoshida T, Kutsuzawa K, et al. Integrated cognitive behavioral therapy for chronic pain: An open-labeled prospective single-arm trial. Medicine (Baltimore) 2021 Feb 12;100(6):e23859 [FREE Full text] [doi: 10.1097/MD.0000000000023859] [Medline: 33578513]
31. John D Otis. Managing Chronic Pain: A Cognitive-Behavioral Therapy Approach Therapist Guide. USA: Oxford University Press; Sep 24, 2007.

32. Marta IER, Baldan SS, Berton AF, Pavam M, da Silva MJP. [The effectiveness of therapeutic touch on pain, depression and sleep in patients with chronic pain: clinical trial]. Rev Esc Enferm USP 2010 Dec;44(4):1100-1106 [FREE Full text] [doi: 10.1590/s0080-62342010000400035] [Medline: 21337796]
33. Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Pers Soc Psychol 1993 Jul;65(1):45-55. [doi: 10.1037//0022-3514.65.1.45] [Medline: 8355141]
34. Cocks M, Moulton C, Luu S, Cil T. What surgeons can learn from athletes: mental practice in sports and surgery. J Surg Educ 2014;71(2):262-269. [doi: 10.1016/j.jsurg.2013.07.002] [Medline: 24602719]
35. García Carrasco D, Aboitiz Cantalapiedra J. Effectiveness of motor imagery or mental practice in functional recovery after stroke: a systematic review. Neurologia 2016;31(1):43-52 [FREE Full text] [doi: 10.1016/j.nrl.2013.02.003] [Medline: 23601759]
36. Tsao JW, Finn SB, Miller ME. Reversal of phantom pain and hand-to-face remapping after brachial plexus avulsion. Ann Clin Transl Neurol 2016 Jun;3(6):463-464 [FREE Full text] [doi: 10.1002/acn3.316] [Medline: 27547774]
37. Flor H, Nikolajsen L, Staehelin Jensen T. Phantom limb pain: a case of maladaptive CNS plasticity? Nat Rev Neurosci 2006 Nov;7(11):873-881. [doi: 10.1038/nrn1991] [Medline: 17053811]
38. Ramachandran VS, Rogers-Ramachandran D, Cobb S. Touching the phantom limb. Nature 1995 Oct 12;377(6549):489-490. [doi: 10.1038/377489a0] [Medline: 7566144]
39. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]
40. Farrar JT, Troxel AB, Stott C, Duncombe P, Jensen MP. Validity, reliability, and clinical importance of change in a 0-10 numeric rating scale measure of spasticity: a post hoc analysis of a randomized, double-blind, placebo-controlled trial. Clin Ther 2008 May;30(5):974-985. [doi: 10.1016/j.clinthera.2008.05.011] [Medline: 18555944]
41. Uki J, Mendoza T, Cleeland CS, Nakamura Y, Takeda F. A brief cancer pain assessment tool in Japanese: the utility of the Japanese Brief Pain Inventory--BPI-J. J Pain Symptom Manage 1998 Dec;16(6):364-373 [FREE Full text] [doi: 10.1016/s0885-3924(98)00098-0] [Medline: 9879161]
42. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532 [FREE Full text] [doi: 10.1037/1040-3590.7.4.524]
43. Sullivan M. The Pain Catastrophizing Scale: User Manual. URL: https://aspecthealth.ca/wp-content/uploads/2017/03/PCSManual_English1.pdf [accessed 2021-11-01]
44. Yamashiro K, Arimura T, Iwaki R, Jensen MP, Kubo C, Hosoi M. A multidimensional measure of pain interference: reliability and validity of the pain disability assessment scale. Clin J Pain 2011 May;27(4):338-343. [doi: 10.1097/AJP.0b013e318204858a] [Medline: 21178590]
45. Kojima M, Nagaya T, Tokudome S, Furukawa T. Development of Beck Depression Inventory-II (BDI-II). J Epidemiol . (in Japanese) 2002;12(1):179 [FREE Full text]
46. Wang Y, Gorenstein C. Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. Braz J Psychiatry 2013;35(4):416-431 [FREE Full text] [doi: 10.1590/1516-4446-2012-1048] [Medline: 24402217]
47. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
48. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
49. King JT, Tsevat J, Roberts MS. Measuring preference-based quality of life using the EuroQol EQ-5D in patients with cerebral aneurysms. Neurosurgery 2009 Sep;65(3):565-72; discussion 572. [doi: 10.1227/01.NEU.0000350980.01519.D8] [Medline: 19687702]
50. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
51. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002 Jun;11(4):341-353. [doi: 10.1002/hec.673] [Medline: 12007165]
52. Shiroiwa T, Sung Y, Fukuda T, Lang H, Bae S, Tsutani K. International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? Health Econ 2010 Apr;19(4):422-437. [doi: 10.1002/hec.1481] [Medline: 19382128]
53. Garg S, Garg D, Turin TC, Chowdhury MFU. Web-Based Interventions for Chronic Back Pain: A Systematic Review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
54. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
55. Cheng JOS, Cheng S. Effectiveness of physical and cognitive-behavioural intervention programmes for chronic musculoskeletal pain in adults: A systematic review and meta-analysis of randomised controlled trials. PLoS One 2019;14(10):e0223367 [FREE Full text] [doi: 10.1371/journal.pone.0223367] [Medline: 31600270]

56. Boersma K, Södermark M, Hesser H, Flink IK, Gerdle B, Linton SJ. Efficacy of a transdiagnostic emotion-focused exposure treatment for chronic pain patients with comorbid anxiety and depression: a randomized controlled trial. Pain 2019 Aug;160(8):1708-1718 [FREE Full text] [doi: 10.1097/j.pain.0000000000001575] [Medline: 31335641]

Abbreviations

ANCOVA: analysis of covariance
BPI: Brief Pain Inventory
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EQ-5D-5L: European quality of life 5-dimensions 5-level
GAD-7: Generalized Anxiety Disorder scale-7 items
ICER: incremental cost-effectiveness ratio
NRS: numerical rating scale
PCS: Pain Catastrophizing Scale
PDAS: Pain Disability Assessment Scale
PHQ-9: Patient Health Questionnaire-9 items
QALY: quality-adjusted life year
RCT: randomized controlled trial
TAU: treatment as usual
vCBT: videoconference-based cognitive behavioral therapy
WTP: willingness to pay

Edited by R Kukafka, G Eysenbach; submitted 03.06.21; peer-reviewed by A Bailey, D Sengupta; comments to author 08.08.21; revised version received 01.09.21; accepted 05.10.21; published 22.11.21.

Please cite as:
Taguchi K, Numata N, Takanashi R, Takemura R, Yoshida T, Kutsuzawa K, Yoshimura K, Nozaki-Taguchi N, Ohtori S, Shimizu E Clinical Effectiveness and Cost-effectiveness of Videoconference-Based Integrated Cognitive Behavioral Therapy for Chronic Pain: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e30690
URL: https://www.jmir.org/2021/11/e30690
doi:10.2196/30690
PMID:34813489

Original Paper

Effects of Internet-Based Cognitive Behavioral Therapy for Harmful Alcohol Use and Alcohol Dependence as Self-help or With Therapist Guidance: Three-Armed Randomized Trial

Magnus Johansson[1,2], PhD; Anne H Berman[2,3], PhD; Kristina Sinadinovic[2], PhD; Philip Lindner[2], PhD; Ulric Hermansson[2], PhD; Sven Andréasson[1], PhD

[1]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

[2]Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden

[3]Department of Psychology, Uppsala University, Uppsala, Sweden

Corresponding Author:
Magnus Johansson, PhD
Department of Global Public Health
Karolinska Institutet
Solnavägen 1E
Stockholm, 11365
Sweden
Phone: 46 727249971
Email: magnus.johansson.1@ki.se

Abstract

Background: Alcohol use is a major contributor to health loss. Many persons with harmful use or alcohol dependence do not obtain treatment because of limited availability or stigma. They may use internet-based interventions as an alternative way of obtaining support. Internet-based interventions have previously been shown to be effective in reducing alcohol consumption in studies that included hazardous use; however, few studies have been conducted with a specific focus on harmful use or alcohol dependence. The importance of therapist guidance in internet-based cognitive behavioral therapy (ICBT) programs is still unclear.

Objective: This trial aims to investigate the effects of a web-based alcohol program with or without therapist guidance among anonymous adult help-seekers.

Methods: A three-armed randomized controlled trial was conducted to compare therapist-guided ICBT and self-help ICBT with an information-only control condition. Swedish-speaking adult internet users with alcohol dependence (3 or more *International Classification of Diseases, Tenth Revision* criteria) or harmful alcohol use (alcohol use disorder identification test>15) were included in the study. Participants in the therapist-guided ICBT and self-help ICBT groups had 12-week access to a program consisting of 5 main modules, as well as a drinking calendar with automatic feedback. Guidance was given by experienced therapists trained in motivational interviewing. The primary outcome measure was weekly alcohol consumption in standard drinks (12 g of ethanol). Secondary outcomes were alcohol-related problems measured using the total alcohol use disorder identification test-score, diagnostic criteria for alcohol dependence and alcohol use disorder, depression, anxiety, health, readiness to change, and access to other treatments or support. Follow-up was conducted 3 (posttreatment) and 6 months after recruitment.

Results: During the recruitment period, from March 2015 to March 2017, 1169 participants were included. Participants had a mean age of 45 (SD 13) years, and 56.72% (663/1169) were women. At the 3-month follow-up, the therapist-guided ICBT and control groups differed significantly in weekly alcohol consumption (−3.84, 95% Cl −6.53 to −1.16; t_{417}=2.81; *P*=.005; Cohen *d*=0.27). No significant differences were found in weekly alcohol consumption between the self-help ICBT group and the therapist-guided ICBT at 3 months, between the self-help ICBT and the control group at 3 months, or between any of the groups at the 6-month follow-up. A limitation of the study was the large number of participants who were completely lost to follow-up (477/1169, 40.8%).

Conclusions: In this study, a therapist-guided ICBT program was not found to be more effective than the same program in a self-help ICBT version for reducing alcohol consumption or other alcohol-related outcomes. In the short run, therapist-guided ICBT was more effective than information. Only some internet help-seekers may need a multisession program and therapist guidance to change their drinking when they use internet-based interventions.

Trial Registration: ClinicalTrials.gov NCT02377726; https://clinicaltrials.gov/ct2/show/NCT02377726

(J Med Internet Res 2021;23(11):e29666) doi:10.2196/29666

KEYWORDS

alcohol dependence; alcohol use disorders; internet-based interventions; internet-based cognitive behavioral therapy; ICBT; cognitive behavioral therapy; CBT; eHealth; alcohol use; substance abuse; outcomes; help-seeking behavior; internet-based interventions; alcohol dependence; mobile phone

Introduction

Background

Alcohol consumption causes substantial health loss. It is the leading risk factor for both mortality and disability-adjusted life years worldwide among those aged 15-49 years, with 3.8% of female deaths and 12.2% of male deaths attributable to alcohol use [1]. There are dose-response relationships between alcohol consumption and many major diseases [2]. Heavy use over time, as discussed by Rehm et al [3], is responsible for most of the attributable burden of disease, mortality, and social consequences of alcohol. When alcohol use is diagnosed as alcohol dependence (*International Classification of Diseases, Tenth Revision* [*ICD-10*]), it is characterized by impaired control and continued heavy use despite negative consequences [4]. The prevalence of alcohol dependence is estimated to be 2.6% worldwide and 3.7% in Europe [5]. In Sweden, 4% of adults are estimated to fulfill the diagnostic criteria for alcohol dependence [6].

Different interventions are available to reduce an individual's alcohol consumption. Brief interventions, often used in primary care settings, are effective in reducing alcohol consumption [7]. Psychological and pharmacological treatments have also shown effects in terms of reduced alcohol consumption (eg, cognitive behavioral therapy [CBT] vs minimal intervention, g=0.67 [8]) and in terms of abstinence from alcohol (eg, acamprosate vs placebo, g=0.36 [9]). However, there have been problems with the implementation of brief interventions; that is, health professionals have limited time or might be reluctant to address alcohol use [10]. Only approximately 7% of people with substance use disorders are estimated to receive at least minimal treatment [11]. Possible reasons for not seeking treatment are lack of access, shame, stigma, not wanting to change one's alcohol use, or a wish to deal with it by oneself [12,13].

Internet-based interventions can help overcome some of the problems associated with implementation, limited accessibility, and stigma [14]. The internet has developed over the past decades from being an alternative way of finding health-related information to being the common way. Approximately >85% of Swedes use the internet to find information about health or medicine, and approximately 40% do so at least every month [15]. Internet-based interventions for reducing alcohol use have been developed during the past 20 years [16] and include content similar to that of face-to-face interventions, such as personal normative feedback, motivational interviewing (MI), and CBT, intended to motivate the user to reduce their drinking and give them strategies to do so [17]. In a Cochrane review of digital alcohol interventions, including internet-based interventions (37/57, 65% studies), the effect compared with no or minimal interventions was 23 g (95% CI 15 to 30) of less alcohol consumed weekly. According to a recent individual patient data meta-analysis of internet-based alcohol interventions, the effect on alcohol consumption compared with various controls was −22 g less per week (95% CI −8.7 to −34.6) [18]. The same meta-analysis also found that guided internet-based alcohol interventions (with human guidance from health professionals or trained volunteers) are more effective than unguided (fully automated) interventions (−67.8 g alcohol per week, 95% CI −121.1 to −14.5).

However, most of the previous studies on digital and internet-based alcohol interventions have been on brief interventions, such as personal normative feedback, and have been limited to at-risk populations, such as students [17,19]. More extended internet-based alcohol interventions are intended to be used over a number of weeks or sessions and are usually internet CBT (ICBT), for example, based on treatments for alcohol dependence, such as relapse prevention or behavioral self-control training, often combined with principles from MI [19,20]. There are indications that longer, multisession interventions are more effective than shorter or single-session internet-based alcohol interventions [21]. A literature search of multisession internet-based alcohol interventions revealed 14 randomized controlled trials of ICBT aimed at drinkers among the general public with at least hazardous use [22-35] (see Multimedia Appendix 1 for further details). In 5 of these studies [23,25,27,28,35], ICBT as self-help was significantly more effective in reducing alcohol consumption than minimal control interventions or waiting lists. However, several previous large studies did not find a significant difference between self-help ICBT and minimal control [26,32,34]. Therapist-guided ICBT for alcohol was tested in 5 of the 14 previous studies and was found to be more effective than waiting list in 4 studies [22,27,28,35] and unguided self-help in 2 studies [24,28]. However, therapist-guided interventions were not more effective than self-help interventions in the 2 most recent studies on ICBT for alcohol [27,35]. Although many of the participants in previous studies of ICBT programs for alcohol have had alcohol use disorder identification test (AUDIT)–scores indicating a high level of alcohol-related problems, there is a need for studies on internet-based interventions that are aimed specifically at people with harmful use or alcohol dependence [19].

As described above, the effects of ICBT programs, as well as therapist-guided ICBT for alcohol, are still unclear. This could be explained by the fact that many trials have used small sample sizes and included users with different levels of problems (eg, included risky alcohol users), who may change their drinking more easily or to a lesser extent. Most previous studies have also used waiting list control conditions or open (unblinded)

design, possibly making the control groups disappointed or less likely to change [36]. In this study, the sample size was larger than in previous studies that investigated the effect of guided ICBT. The participants were also blinded to the interventions that the other participants received. The purpose of this randomized controlled trial was to investigate the effects of a web-based alcohol program with or without therapist guidance among anonymous adult help-seekers with harmful use or alcohol dependence.

Hypotheses

The hypotheses of the trial were:

1. A therapist-guided ICBT program would lead to a greater reduction in alcohol consumption and alcohol-related problems than information alone.
2. A self-help ICBT program would lead to a greater reduction in alcohol consumption and alcohol-related problems than information alone.
3. A therapist-guided ICBT program would lead to a greater reduction in alcohol consumption and alcohol-related problems than a self-help ICBT program.

Methods

Study Design

In a three-arm randomized controlled trial with a parallel design, participants were randomly assigned to an internet-delivered CBT (ICBT) program as self-help, with therapist guidance or information control in a ratio of 1:1:1 and a block size of 30. The trial was approved by the Stockholm Regional Ethical Review Board (No. 2014/1758-31/2).

Recruitment

Overview

Participants were recruited at the Swedish internet site Alkoholhjälpen [37], an open access website that provides information and a discussion forum for individuals seeking web-based help for alcohol consumption. The site has been publicly accessible since 2007. During the recruitment period of this study, Alkoholhjälpen had approximately 20,000 unique visitors every month and approximately 100 new forum posts every day. All service use was free of charge, and no advertising was allowed on the website. All visitors on Alkoholhjälpen from March 2015 to March 2017 were invited to participate in a study to develop and test different forms of internet-delivered support for changing alcohol habits. Interested users were informed that they would answer a survey and be randomized to one of three different forms of support but were not informed about the specifics. Adult individuals who gave their informed consent were instructed to create a personal account with a unique username and password. They were then directed to a screening page where they were required to give informed consent for participation in the study, answer demographic questions, questions in AUDIT [38], and questions about alcohol dependence (*ICD-10*) criteria.

Inclusion Criteria

Individuals were included if they had harmful use (defined as >15 total score in AUDIT) or alcohol dependence (defined as 3 or more *ICD-10* criteria). Registrants who did not meet either of these criteria were informed that they did not qualify for the study and were invited to use the open parts of the website. To be able to complete the registration, the participants needed to understand written Swedish and be computer literate enough to access and navigate the website via a computer, tablet, or smartphone. Before registering, potential participants were also informed that the interventions were not intended for users who were experiencing withdrawal symptoms, psychosis, schizophrenia, bipolar disorder, or suicidal thoughts.

Baseline Assessment

Eligible participants were asked to complete web-based baseline questionnaires, including primary and secondary measures (*Measures* section). All participants were treated as anonymous users. They were asked to provide an email address and a mobile phone number for notifications and follow-up reminders. Email and phone numbers were neither verified nor used for identification or for any other purposes. Participants whose assessments showed indications of an increased risk of suicide (Montgomery Asberg Depression Rating Scale–Self-rated [MADRS-S] item 9>3) at baseline or follow-up received a message offering additional support with phone numbers to call in case of emergency. No eligible participants were excluded based on baseline assessment. The baseline questionnaires were followed by a survey on why participants chose to use web-based services and their preferences regarding such services. Data from this survey will be presented elsewhere.

Randomization

Participants who completed the baseline measures were randomized according to a fully automated and concealed procedure on the web-based platform. Participants were assigned to one of 3 groups: (1) self-help ICBT: a self-help program; (2) therapist-guided ICBT: a program with web-based guidance from a therapist; or (3) control: information on changing alcohol habits. Participants were blinded to the kind of support received by participants in the other groups.

Intervention Groups

Self-help ICBT

Directly following randomization, the self-help ICBT group was given access to the program. The program was based on self-help materials used in previous studies on the internet and in specialist care [24,34,39,40]. Content and exercises in the program were based on MI [41,42], relapse prevention [43,44], and behavioral self-control [45,46]. The program was divided into 5 main modules, 3 extra problem-solving modules, and 10 fact sheets (refer list of modules in Table 1). The length of the program was approximately 17,000 words in total, with 5500 words in the extra modules and 3000 words in the fact sheets. The module texts were alternated with checklists and open questions that prompted the user to give their view of the content in relation to their own situation. The modules also included videos with examples or expert-interviews. Refer to Figure 1 for example pictures of the program. Automatic reminders with

suggestions on what module to work on were sent at 1-4, 6, and 8 weeks. Users were also encouraged to register alcohol consumption or craving, as well as details of the situation when they drank or experienced a craving. This was done in a private drinking calendar included in the web program, which could be used daily or for a whole week retrospectively. Continual feedback on the users' alcohol consumption was offered through a private statistics page. Here, users could see their average personal consumption in standard drinks weekly, monthly, and in total, as well as the number of days drinking, the number of days sober, and binge drinking occasions. In addition, they could also view a summary of their own risk situations. To allow participants to complete the modules at the recommended pace, with room for some delay, the participants had access to the program for 12 weeks after allocation.

Table 1. Program modules and number and percentage of participants in the therapist-guided and self-help internet-based cognitive behavioral therapy (ICBT) who used each module.

Module[a]	Therapist-guided ICBT (n=386), n (%)	Self-help ICBT (n=391), n (%)
Motivation (including brief feedback on assessment)	272 (70.5)	265 (67.6)
Drinking-goal and self-control	202 (52.3)	172 (43.8)
Behavioral analysis of drinking and risk situations	148 (38.3)	113 (28.8)
General problem solving	102 (26.4)	73 (18.6)
Handling cravings	103 (26.7)	72 (18.4)
Handling feelings	77 (19.9)	46 (11.7)
Drink-refusal skills	65 (16.8)	35 (8.9)
Preventing relapse	63 (16.3)	34 (8.6)

[a]All modules were available to the user from the start. Each module contained general information, audio or video, examples, and exercises. Additional fact sheets included in the program concerned blood alcohol level, anxiety, depression, anger, stress, managing thoughts, relaxation, sleep, leisure activities, and communication.

Figure 1. Program module on a computer and the drinking calendar on a smartphone.

Therapist-Guided ICBT

Participants in the therapist-guided ICBT group had the same access to the same program as the self-help ICBT described above but the drinking calendar and module answers were shared with a therapist. The therapist-guided ICBT group could communicate with the therapist through asynchronous SMS text messages on the intervention website during the 12 weeks of the program. The guidance from the therapist focused on motivating the user to continue using the program and change their drinking. Each time the participant had completed any of the modules, the therapist wrote personal feedback and answered any questions about the program via private comments on the web platform. The feedback highlighted parts of what the participant had stated in the exercises included in each module, which were important from an MI or CBT perspective. Users

who did not use or stopped using the program for several weeks were reminded by the therapist 2 times through personal messages on the website (with notification on mail). The 3 therapists involved had several years of experience on the Swedish alcohol helpline [47] and had all reached an approved level in phone-based MI before entering the study. They had been trained in ICBT and received regular supervision from the first author, who is a trained therapist with several years of experience with CBT and ICBT programs.

Control

The control group was given access to text-only information on changing their alcohol habits based on the text Alcohol and you [41]. The information material was equivalent to 5 pages of printed text.

All 3 groups also had access to the discussion forum on the website as well as facts on alcohol and health and information about how to find additional support within the health care or social welfare system. Communication between the server hosting the intervention and the participant was encrypted and protected with an individual login name and password.

Measures

Overview

Follow-up was conducted 3 (after treatment), 6, 12, and 24 months after recruitment. In this paper, the results from the 3- and 6-month follow-ups are presented. Primary and secondary outcomes were assessed at all time points.

Primary Outcome

The primary outcome was the difference between the groups in alcohol consumption and mean weekly standard drinks. The number of standard drinks each of the 7 days in the preceding week was self-reported using the timeline follow-back (TLFB) method [48]. One standard drink contains 12 g of pure alcohol, according to the Swedish definition. The TLFB has been shown to be a valid and reliable procedure to document recent drinking when administered via the internet [49] and in a 7-day version [50].

Secondary Outcomes

Alcohol-related outcomes were assessed with a number of different instruments. AUDIT [38] is a 10-item instrument that covers both alcohol consumption and problems and has been validated in Swedish and via the computer [51,52]. The total AUDIT score was used as a continuous measure of alcohol-related problems. For description at baseline as well as for assessment of nonimproved and deteriorated participants at the 6-month follow-up, the AUDIT score was categorized into 4 zones as follows: I nonproblematic, 0-6p; II hazardous use, 7-15p; III harmful use, 16-19p; and IV probable dependence, 20-40p. The sum of the 3 first items (AUDIT-C) was also used to assess alcohol consumption [53]. Alcohol dependence was assessed by the number of self-reported alcohol dependence criteria during the past year according to *ICD-10* [4]. Alcohol use disorder was assessed by the number of self-rated alcohol use disorder criteria during the past year according to the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) (*DSM-5*) [54]. Measures of both alcohol dependence and alcohol use disorder diagnostic criteria were included to facilitate comparisons between the study population and diverse community and clinical populations reported in the research literature, where problem severity level can be assessed according to both *ICD-10* and *DSM-5*.

At the 3- and 6-month follow-ups, the time frame for AUDIT, *ICD-10*, and *DSM-5* was changed from 12 months to 3 months. The number of nondrinking days, number of binge drinking days (defined as ≥3 drinks for women and ≥4 drinks for men), the average number of drinks on drinking days, and low-risk consumption at follow-up (≤14 drinks per week for men and ≤9 drinks per week for women and no binge drinking) were also assessed using the TLFB. Health-related quality of life was assessed using the EuroQol-5 dimensions (EQ-5D-5L). An index score was calculated with Crosswalk value sets, using the United Kingdom as a reference [55,56]. Symptoms of depression were measured using the total score of MADRS-S [57,58]. Symptoms of anxiety were measured using the total score of the Generalized Anxiety Disorder Assessment-7 items (GAD-7) [59,60]. *ICD-10*, *DSM-5*, MADRS-S, and GAD-7-scores were categorized for description at baseline (see Multimedia Appendix 2 for details). Use of other support was assessed by 4 questions covering who and where participants talked to someone about their alcohol problems and which medication or which other internet resources they had used regarding alcohol.

Additional Measures

Readiness to change was measured at all time points with 2 visual analog scales [61], where users responded on a scale of 0-10 from “I am not ready to reduce/quit my drinking” (0) to “I am very much ready to reduce/quit my drinking” (10). Working alliance was measured 3 and 6 weeks after randomization when all participants were invited via one email and message to log in to the website and answer the Session Rating Scale [62] regarding the use of the website and the intervention that they had received. After each module in the program, users in the therapist-guided ICBT and self-help ICBT groups could rate, on a scale of 0 to 5 stars, how helpful they found the module to be. All uses of the intervention were logged for each user. Participants who completed 4 or more modules in the program were regarded as treatment completers.

Follow-up

Follow-up was conducted at 3 and 6 months after recruitment. Based on previous web-based studies, high attrition from follow-up could be expected [63]. Reminders and incentives [64] were used to prevent attrition without affecting the intended target group by forcing participants to have personal contact or identify themselves. At follow-up, participants were emailed a link or redirected when logging in on the intervention website to the follow-up questionnaires. The email included information that all participants who completed follow-ups would have a 0.8% (1/120) chance of receiving a free iPad. The same questionnaires, with all primary and secondary outcomes used at baseline, were also used at follow-up and adjusted for the time since the last assessment (3 months). Participants who did not respond to this initial request received up to 5 automated email reminders, a manual email reminder, and a mobile SMS

text message. Additional follow-ups at 12 and 24 months after recruitment have recently been completed but have not yet been analyzed.

Sample Size

Sample size was determined a priori using an effect-size estimate. We aimed to detect a Cohen *d*=0.2 in 2-group comparisons using 2-tailed *t* tests at follow-ups, which based on SDs from a study of alcohol treatment in primary care [40], equated to a between-group difference of Δmean=3.7 drinks. With α=.05 and 80% power, n=394 per group was required for the desired effect size, totaling n=1182. However, to allow analyses of observed data only, assuming 50% missing data at follow-up, the enrollment goal was increased to n=2400.

Analytic Plan and Statistical Procedure

All statistical analyses were two-sided tests and, unless otherwise specified, used a significance level of *P*<.05. Factorial analysis of variances was used to test differences in baseline measures between users who were retained and those who were lost to follow-up, including interactions between groups and lost to follow-up (at either 3 months or 6 months). Differences in categorical measures were analyzed using the chi-square test. All tests were performed using SPSS version 25 (IBM Corporation).

In accordance with the original protocol, differences in observed means at each follow-up were analyzed with *t* tests under the "missing at random" assumption; however, significant contrasts were supplemented with tipping point sensitivity analyses that systematically imputed missing data at a group level across a range of plausible mean values in the 2 nonrespondent groups (with the same SD) [65,66]. A custom R function was developed that, per contrast, created a matrix of all possible combinations of group-level imputed means (in steps of 0.1 and within the plausible range) of the 2 nonrespondent groups, calculated the new whole-group means and SDs, performed a *t* test using this summary data, and saved the *P* value. This allowed us to estimate the conditions under which (means among nonrespondents) the contrast would no longer be significant (tipping point), thereby testing the appropriateness of the "missing at random" assumption. Two-group *t* tests were corrected for multiple comparisons by considering *P*<.02 as significant, corresponding to the Bonferroni adjustment.

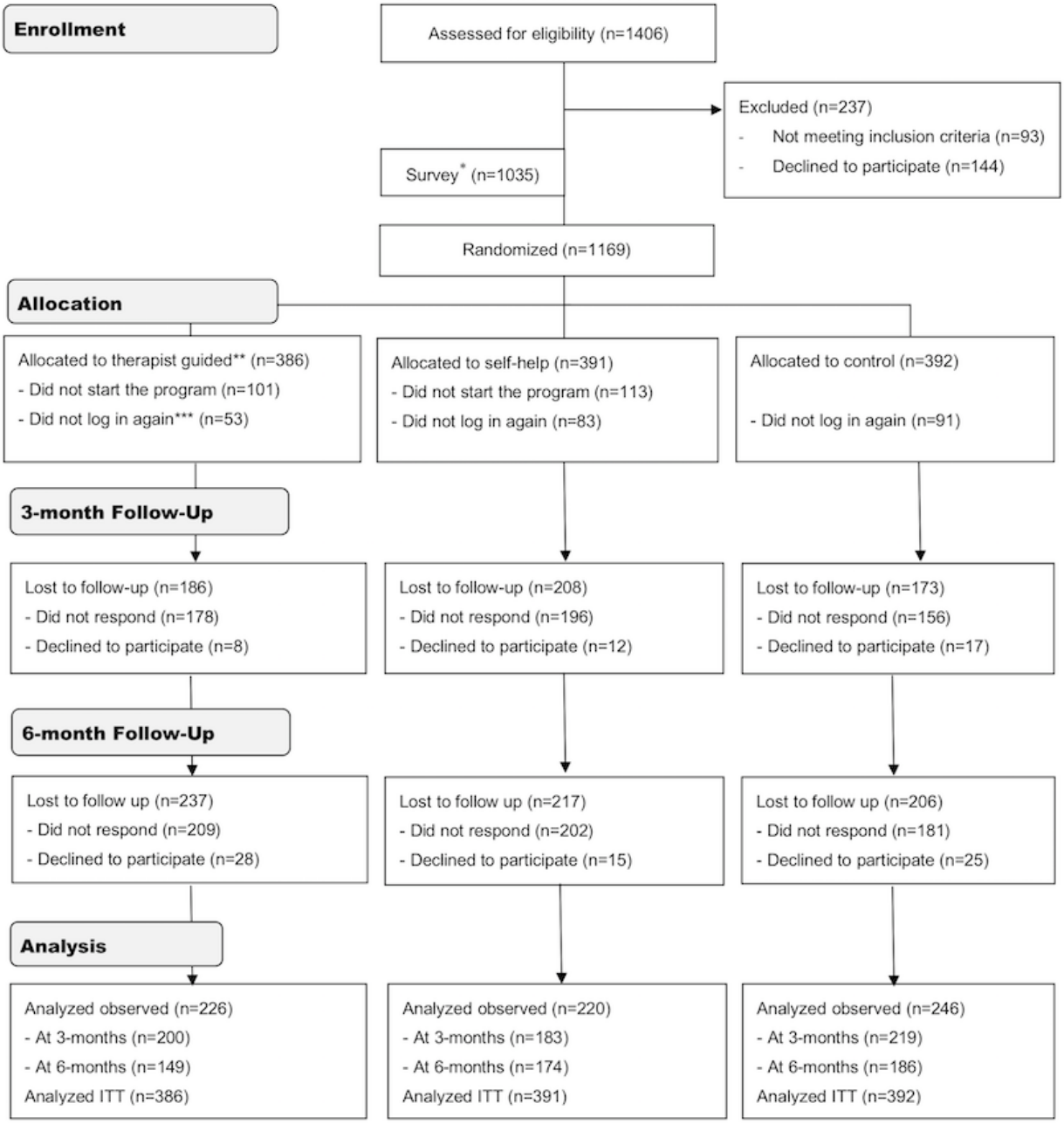
Before analyzing the data, the decision was made to supplement the original analytic protocol with mixed-effects modeling that would be fully compliant with the intention-to-treat principle and better equipped to handle the presumed high degree of missing data [63]. By modeling data at both group (fixed) and individual (random) levels, mixed models are well-suited for data from repeated observations (modeling clustering of data at an individual level) [67], and maximum likelihood estimation is used to handle missing data [68]. Analyzing outcomes with (generalized) mixed-effects models also allowed the use of family functions that were more appropriate for the distribution of the outcome. Weekly alcohol consumption (as well as other outcomes based on TLFB) was analyzed using generalized linear mixed models [69] with a negative binomial distribution and log link to avoid overestimation of effects [70]. AUDIT, *ICD-10*, *DSM-5*, GAD-7, MADRS-S, readiness, and EQ-5D-5L were analyzed with linear mixed models. Mixed-effects modeling was performed using SPSS 25. First, a random intercept model was specified to calculate an interclass correlation score [71]. After visually inspecting the average and individual growth curves, an unconditional model was specified by adding both linear and quadratic time (time^2) as predictors. A conditional growth model was then specified by adding the self-help ICBT and therapist-guided ICBT groups as dummy-coded variables to be compared against the control (reference), together with the time×group and time^2×group interactions. Different covariance structures for random effects and errors were tested, and a likelihood ratio test was used to assess which model best fitted the data [67]. There was a significant dependency among the observations (intraclass correlation=0.62) as well as significant individual variability in the initial level (intercept *P*<.001) and rate of change (slope *P*<.001) in the primary outcome.

Results

Participants

A total of 1169 participants were randomly allocated to the 3 study arms (refer to Figure 2 for the flow chart). This was lower than the target sample size required to adjust analyses on observed data only for estimated attrition at follow-up but only negligibly smaller than in the raw power calculation with estimated missing data (n=1182). Nonetheless, recruitment ceased after the prespecified 24 months recruitment window for funding reasons. Individuals who declined participation after the screening did not differ from those included in the AUDIT (mean 22.1, SD 5.6 vs mean 22.6, SD 6.5; t_{1308}=1.00; *P*=.32) or *ICD-10* (mean 4.2, SD 1.3 vs mean 4.2, SD 1.2; t_{1308}=0.10; *P*=.92) scores. The randomized participants had a mean age of 45 years (SD 13), and 56.72% (663/1169) were women. During the past year 5.05% (59/1169) had talked to someone in specialized care and 18.99% (222/1169) to a professional about their alcohol use. Participants indicated significantly higher mean readiness to reduce their drinking compared with mean readiness to quit drinking (mean 8.8, SD 1.9 vs mean 5.7, SD 3.6; t_{1168}=32.4; *P*<.001). The full demographic and clinical variables at baseline are shown in Multimedia Appendix 2.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. *All included participants were asked to complete a survey on why they had chosen to use web-based services and their preferences regarding such services. **Overall, 390 users were randomized to the therapist group, but 4 of them (due to a temporary technical error) never completed the baseline assessment and were never allocated to the intervention. ***Users who did not log in a second time after allocation to intervention. ICBT: internet-based cognitive behavioral therapy; ITT: intention-to-treat.



Loss to Follow-up

Attrition was 49% at the 3-month and 66% at the 6-month follow-up. The number of participants who completed at least one of the follow-ups at 3 or 6 months was 692 (59.2%). There were no significant differences in the number of participants in each group who completed at least one of the follow-ups (X^2_2=3.5; *P*=.17). Comparisons between those lost to follow-up and those who completed at least one follow-up showed greater baseline drinking, as well as greater severity on most baseline measures, among those who were lost to follow-up (Table 2). However, no significant interactions between group assignment and being lost to follow-up were found for baseline outcome variables.

Table 2. Differences in baseline variables between participants who were lost to follow-up (both assessment) and those who completed at least one assessment (N=1169).

Characteristics	Therapist-guided ICBT[a] (n=386), mean (SD)		Self-help ICBT (n=391), mean (SD)		Control (n=392), mean (SD)		Comparison			
	Retained (n=226)	Lost (n=160)	Retained (n=220)	Lost (n=171)	Retained (n=246)	Lost (n=146)	Lost vs retained		Group×lost	
							F test (*df*)	*P* value	*F* test (*df*)	*P* value
Weekly drinks	23.7 (16.7)	25.9 (16.1)	24.4 (17.5)	28.3 (17.4)	24.4 (17.5)	28.3 (17.4)	11.99 (1)	<.001	0.43 (2)	.65
AUDIT[b]	21.6 (5.4)	22.9 (5.9)	21.9 (5.6)	22.6 (5.5)	21.9 (5.6)	22.6 (5.5)	6.89 (1)	.009	0.42 (2)	.66
AUDIT-C[c]	8.1 (1.8)	8.6 (1.5)	8.1 (1.8)	8.5 (1.7)	8.1 (1.8)	8.5 (1.7)	15.07 (1)	<.001	0.04 (2)	.96
ICD-10 [d]	4.2 (1.4)	4.3 (1.3)	4.15 (1.4)	4.5 (1.3)	4.2 (1.4)	4.5 (1.3)	4.39 (1)	.04	1.36 (2)	.26
DSM-5 [e]	7.0 (2.4)	7.3 (2.3)	7.07 (2.3)	7.5 (2.1)	7.1 (2.3)	7.5 (2.1)	5.48 (1)	.02	0.14 (2)	.87
MADRS-S[f]	18.9 (8.7)	19.1 (10.0)	17.3 (8.9)	18.8 (8.8)	17.3 (8.9)	18.8 (8.8)	3.87 (1)	.049	0.62 (2)	.54
GAD-7[g]	8.3 (5.3)	9.4 (5.7)	7.7 (5.4)	8.8 (5.2)	7.7 (5.4)	8.8 (5.2)	12.82 (1)	<.001	0.01 (2)	.99
EQ-5D[h]	0.72 (0.19)	0.69 (0.24)	0.73 (0.19)	0.72 (0.18)	0.73 (0.19)	0.72 (0.18)	7.37 (1)	.007	1.05 (2)	.35
Nondrinking days	2.9 (1.97)	2.7 (2.08)	2.85 (2.16)	2.6 (2.08)	2.85 (2.16)	2.6 (2.08)	3.61 (1)	.06	0.01 (2)	.99
Binge drinking days	2.7 (1.9)	3.0 (1.8)	2.8 (2.1)	3.3 (2.1)	2.8 (2.1)	3.3 (2.1)	10.59 (1)	<.001	0.30 (2)	.74
Drinks per drinking day	5.1 (3.7)	5.2 (4.1)	4.8 (3.8)	4.9 (4.2)	4.8 (3.8)	4.9 (4.2)	2.43 (1)	.12	0.86 (2)	.42

[a]ICBT: internet-based cognitive behavioral therapy.

[b]AUDIT: alcohol use disorder identification test.

[c]AUDIT-C: alcohol use disorder identification test consumption questions.

[d]*ICD-10*: *International Classification of Diseases, Tenth Revision*.

[e]*DSM-5*: *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition).

[f]MADRS-S: Montgomery Asberg Depression Rating Scale–Self-rated.

[g]GAD-7: Generalized Anxiety Disorder Assessment–7 items.

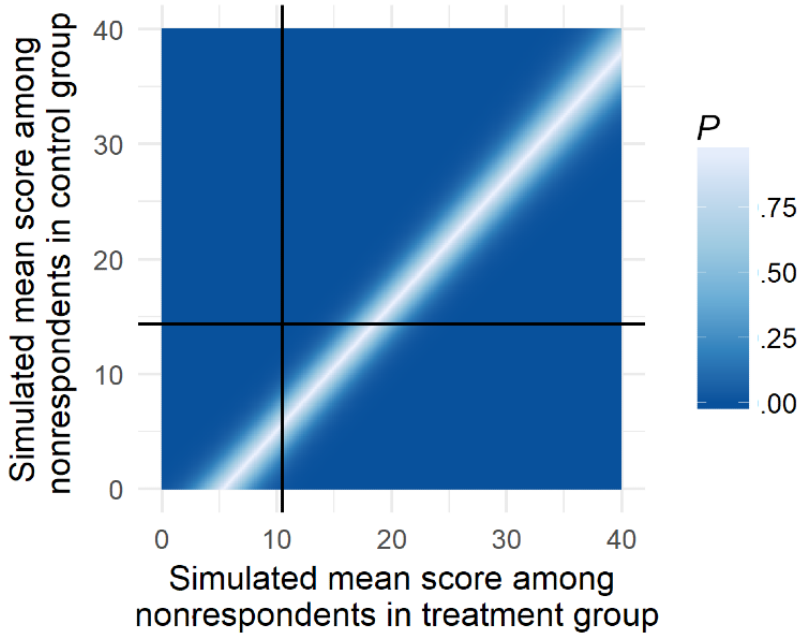
[h]EQ-5D: EuroQol-5 dimensions.

Differences in Observed Means

Outcomes at 3 Months (Posttreatment)

The therapist-guided ICBT group had significantly lower weekly alcohol consumption than the control group at 3 months (mean difference −3.84, 95% Cl −6.53 to −1.16; Cohen *d*=0.27). A tipping point sensitivity analysis revealed that missing data in the control group would need to have a mean of 11.0 to render this contrast insignificant (assuming missing at random in the therapist-guided ICBT group), or that the missing data in the therapist-guided ICBT group would need to have a mean of 13.6 (assuming missing at random in the control group). Refer to Figure 3 for the *P* value heat map for each of the possible combinations of imputed means among nonrespondents in the 2 groups. No significant differences in weekly alcohol consumption were found between the self-help ICBT group and the control group (mean difference −2.41, 95% Cl −5.53 to 0.71) or between the therapist-guided ICBT group and the self-help ICBT group (mean difference −1.43, 95 CI −4.26 to 1.40).

Figure 3. Tipping point analysis of significant contrast in primary outcome at 3 months.



At 3 months (posttreatment) there were significant differences between the therapist-guided ICBT group and the control group in the secondary outcomes AUDIT (mean difference −2.91, 95% Cl −4.33 to −1.50; Cohen *d*=0.39), AUDIT-C (mean difference −0.77, 95% CI −1.34 to −.20; Cohen *d*=0.26), *DSM-5* (mean difference −0.76, 95% CI −1.34 to −0.17; Cohen *d*=0.25), and *ICD-10* (mean difference −0.47, 95% Cl −0.82 to −0.14; Cohen *d*=0.26). A difference was also found between the self-help ICBT and control groups on the AUDIT (mean difference −1.95, 95% Cl −3.44 to −46; Cohen *d*=0.26) at 3 months. No significant differences in secondary outcomes between the therapist-guided ICBT and self-help ICBT were found at 3 months. See Multimedia Appendix 3 for details and Multimedia Appendix 4 for the tipping point analyses of secondary outcomes. Of the participants who completed the 3-month follow-up, 42.9% (258/602) reported low-risk alcohol consumption (14 or fewer drinks per week for men and 9 or fewer drinks per week for women and no binge drinking) at 3 months, with no significant differences between groups (X^2_2=1.9; *P*=.38).

Outcomes at 6 Months

There were no significant differences in weekly alcohol consumption among any of the groups at 6 months. The difference in means between the therapist-guided ICBT and control group was −0.60 (95 CI −3.70 to 2.50), between self-help ICBT and control group −0.45 (95% Cl −3.87 to 2.96), and between the therapist-guided ICBT and the self-help ICBT −0.15 (95% Cl −3.70 to 3.41). No significant differences in secondary outcomes were found among any of the groups at 6 months. Low-risk alcohol consumption was reported by 42.6% (217/509) of participants at 6 months, with no significant differences between the groups (X^2_2=0.2; *P*=.92). Participants reported no adverse events because of the intervention. Among participants followed up at 6 months, 3.3% (17/508) had changed their alcohol use to a more severe category according to AUDIT, and 19.5% (99/508) remained in the highest AUDIT category.

Post Hoc Intention-To-Treat Mixed Models Analysis

The mixed-model intention-to-treat analysis showed a significantly larger decrease in weekly alcohol consumption over time in the therapist-guided ICBT (time×therapist; *P*=.02) compared with the control but not in the self-help ICBT (time×self-help; *P*=.09) compared with the control (refer models in Table 3). A model comparing only the participants in the therapist-guided ICBT and self-help ICBT did not reveal any significant effect of group×time (*P*=.57). There was a significant decrease in weekly alcohol consumption over time for participants in all 3 groups according to the estimate of time in the unconditional model (*P*<.001).

Table 3. Post hoc mixed models of weekly alcohol consumption, alcohol-related problems, alcohol use disorder, and alcohol dependence[a].

Characteristics	Weekly drinks[b]		AUDIT[c,d]		AUDIT-C[c,e]		*ICD-10* [c,f]		*DSM-5* [c,g]	
	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value	B (SE)	*P* value
Intercept	3.06 (0.04)	>.001	22.03 (75.44)	>.001	8.29 (82.50)	>.001	4.21 (58.59)	>.001	7.14 (58.38)	>.001
Time	−0.92 (0.14)	>.001	−10.40 (−13.27)	>.001	−4.00 (−13.40)	>.001	−1.42 (−7.35)	>.001	−2.68 (−8.29)	>.001
Time^2	0.17 (0.07)	.02	3.32 (8.47)	>.001	1.32 (8.84)	>.001	0.47 (4.82)	>.001	0.90 (5.55)	>.001
Therapist-guided ICBT[h]	−0.03 (0.05)	.55	0.07 (0.16)	.87	0.03 (0.20)	.84	0.03 (0.29)	.77	0.03 (0.16)	.87
Self-help ICBT	0.00 (0.05)	.94	0.20 (0.49)	.62	0.01 (0.04)	.97	0.09 (0.86)	.39	0.12 (0.67)	.50
Therapist-guided ICBT×time	−0.48 (0.20)	.02	−4.51 (−3.99)	>.001	−1.40 (−3.26)	.001	−0.63 (−2.27)	.02	−0.98 (−2.10)	.04
Therapist-guided ICBT×time^2	0.23 (0.11)	.03	1.84 (3.24)	.001	0.62 (2.89)	.004	0.22 (1.57)	.12	0.34 (1.46)	.14
Self-help ICBT×time	−0.35 (0.21)	.09	−3.02 (−2.60)	.009	−0.59 (−1.35)	.18	−0.35 (−1.24)	.22	−0.43 (−0.90)	.37
Self-help ICBT×time^2	0.17 (0.11)	.13	1.18 (2.01)	.04	0.25 (1.14)	.26	0.10 (0.69)	.49	0.11 (0.44)	.66

[a]Time was coded in 3-month-periods (0, 1 and 2). A model with quadratic time (time^2) was chosen since it fitted the data better than a model with linear time only. Reference group was control.

[b]Generalized linear mixed-model. Neg-binominal distribution, dispersion coefficient: 0.944. Covariance structure for random effects Variance component and for repeated effects Diagonal.

[c]Linear mixed-model. Covariance structure for random effects Variance component and for repeated effects First-Order Autoregressive.

[d]AUDIT: alcohol use disorder identification test.

[e]AUDIT-C: alcohol use disorder identification test consumption questions.

[f]*ICD-10*: *International Classification of Diseases, Tenth Revision*.

[g]*DSM-5*: *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition).

[h]ICBT: internet-based cognitive behavioral therapy.

According to the mixed-model analysis of secondary outcomes, a significantly larger decrease in AUDIT over time was found for both therapist-guided ICBT and self-help ICBT compared with the control group, as shown by therapist-guided ICBT×time (*P*<.001) and self-help ICBT×time (*P*=.03). There were also significant therapist-guided ICBT×time group effects for the AUDIT-C (*P*=.01), *ICD-10* (*P*=.02), and *DSM-5* (*P*=.04) diagnostic criteria. No other significant time×group effects were found in the mixed-model analysis of secondary outcomes. Over time, there was a significant decrease among all participants on the AUDIT, AUDIT-C, *ICD-10*, *DSM-5*, MADRS-S, GAD-7, binge drinking days, and drinks on drinking days as well as a significant increase in EQ-5D-5L and nondrinking days. See Multimedia Appendix 5 for additional models.

Changes at follow-ups showed strong correlations between similar outcome variables, such as alcohol dependence and AUD, GAD7, and MADRS-S or weekly drinking and binge drinking days but only moderate or weak correlations between alcohol-related variables and other outcomes. Refer to Multimedia Appendix 6 for further details.

Intervention Use and Rating

The number of modules completed by the therapist-guided ICBT (mean 3.3, SD 3.5) was significantly higher (t_{775}=2.9; *P*=.004) compared with the self-help ICBT (mean 2.6, SD 3.2); however, there was no significant difference in the number of calendar entries (therapist-guided ICBT: mean 39, SD 61 and self-help ICBT: mean 37; SD 102; t_{776}=0.60; *P*=.58). In the therapist-guided ICBT, 39.9% (154/386) were treatment completers, and in the self-help ICBT, 30.4% (119/386) (X^2_1=6.5; *P*=.01). In the therapist-guided ICBT, 58% (224/386) sent at least one message to their therapists. They sent a mean 4.7 (SD 4.7) messages and received a mean 6.0 (SD 4.1) from their therapist. Refer to Table 1 for the details of program use. The number of participants who used the discussion forum was higher in the control group compared with the therapist-guided ICBT group (107/386, 27.7% vs 71/391, 18.2%; X^2_1=8.7; *P*=.003). However, there were no significant differences between self-help ICBT (88/392, 22.4%) and the therapist-guided ICBT or the control group in forum use. Of those who participated in at least one follow-up, 47% (325/692) answered that they had talked to a professional about their alcohol use since entering the study, there were no significant differences between the

groups. Participants gave a significantly higher rating of the working alliance with the intervention (Session Rating Scale) in the therapist-guided ICBT (mean 27, SD 11; t_{269}; $P<.001$) and self-help ICBT (mean 29, SD 10; 185/392, 47.1%; t_{305}; $P<.001$) compared with the information control group (mean 19, SD 12). The mean rating of program modules were 3.8 (SD 1.0) on a scale 0-5, with no difference in rating between the therapist-guided ICBT and the self-help ICBT ($t_{731}=0.69$; $P=.49$).

Discussion

Principal Findings

The aim of this web-based randomized controlled trial was to investigate the effect of a program for harmful alcohol use and alcohol dependence, delivered as self-help ICBT or therapist-guided ICBT. The results only partly confirmed the first hypothesis. Participants randomized to therapist-guided ICBT reduced their weekly alcohol consumption as well as alcohol-related problems (measured with AUDIT) and signs of alcohol use disorder significantly more than participants in the control group at the 3-month follow-up. These findings are in line with the results of previous studies on therapist-guided ICBT [22,24,27] but with smaller differences between the groups. The results did not confirm the second hypothesis. Self-help ICBT was not more effective than the control condition in changing alcohol consumption. This is in line with the results of the first study of Alkoholhjälpen [34] and 2 other large studies of publicly available services [26,31]. However, the self-help ICBT group did change their alcohol-related problems significantly more than the controls at 3 months. No support was found for the third hypothesis. There were no significant differences in changed drinking or other outcomes between therapist-guided ICBT and self-help ICBT. This finding differs from our pilot study [24] and the previous study by Blankers et al [28], in which therapist guidance was significantly more effective than self-help, but in line with recent studies by Boß et al [27] and Sundström et al [35].

A possible factor explaining the difference in results between trials could be the intensity of guidance [72]. Participants receiving therapist-guided interventions completed approximately 60% of the programs in the 3 trials for which data are available [22,24,27]. The study by Postel et al [22], showing the largest effects of a therapist-guided intervention, had a high level of guidance, whereas the study by Boß et al [27] had low intensity, with only 33% of the therapist-guided group using the guidance. In this study, 58% (224/386) of patients in the therapist-guided group used the guidance, which was neither low nor high in intensity compared with the previous studies. The differences between the groups were smaller than those in previous studies. This might be explained by the fact that the participants in this trial were blinded to group allocation, which reduced the risk of being negatively affected by being put in a control group [36,73]. The control group in this study reduced their alcohol consumption by 11 weekly drinks between baseline and the 3-month follow-up compared with, for example, 3 drinks for the waiting list in the study by Postel et al [22]. The difference between groups at 3 months faded at 6 months in this trial. The therapist guidance and program ended after 12 weeks; offering more extended support might have increased the effects at 6-months.

The results of this study also suggest that the decrease in alcohol consumption and related outcomes might result from factors other than the interventions that affected all participants or occurred before the randomization. A significant decrease in alcohol consumption and alcohol-related problems occurred in all 3 study groups in the first 3 months, and this decrease remained stable up to 6 months after inclusion. All study participants were recruited based on their initial harmful use or alcohol dependence. Regression toward the mean [74] could explain some of the decreases in alcohol use and other outcomes in all 3 groups. Participants were recruited from a website about changing alcohol use and could be characterized as help-seekers. Even though they did not know what kind of interventions they would receive, they had already taken steps in the direction of changing their alcohol consumption by signing up to the website. This was also indicated by the high mean readiness to change alcohol consumption (8.8/10). All participants also had to answer a large number of assessment questions about their alcohol, an activity that has been shown to lead to reductions in alcohol consumption [75,76].

All 3 groups in this trial had access to a well-established and moderated discussion forum, which might have affected their alcohol use [77,78]. Significantly more participants in the control group used the forum than in the therapist-guided ICBT group, which might indicate that some participants compensated for the lack of human contact in the intervention by using the forum. Symptoms of depression and anxiety were reduced, and health improved over time in all 3 groups but not more in those who received self-help ICBT or therapist-guided ICBT. This finding differs from findings by Boß et al [27], where a multisession program for risky drinking had a small-size effect on depression, stress, and anxiety at follow-up relative to control.

Generalizability

The generalizability is likely limited to Swedish-speaking people with harmful alcohol use or alcohol dependence seeking help for their drinking on the internet. This study tried to recruit participants that were as similar as possible to the intended target group of the intervention as used in regular service at Alkoholhjälpen. The trial was conducted in the same setting. Information on needed language skills and the limitations of the interventions in helping those with severe psychiatric problems were provided; however, no other criteria were used to exclude participants who fulfilled the criteria set for harmful use or alcohol dependence. There were no differences in alcohol-related problems (AUDIT) or dependence criteria (*ICD-10*) between those who accepted and those who declined participation. The high attrition rate also limits the generalizability of our results.

Strengths

This randomized controlled study of therapist-guided ICBT and self-help ICBT is one of the largest among anonymous internet help-seekers to date. The study reached a large number of people

with harmful use or alcohol dependence, most of whom had not previously received support from specialized care. The number of dependence criteria met by participants was similar to that in a recent Swedish clinical trial in specialized and primary care [40]. The participants in this study were relatively well-educated, full-time employed individuals with stable living arrangements and with equal representation of men and women. This represents most individuals with alcohol dependence [6], a population that is different from those who usually receive treatment for alcohol use disorders [79,80], but that might be reached with internet-based interventions.

Limitations

Despite great efforts to remind and reinforce participants to answer follow-up questions, attrition was high. Participants lost to follow-up showed some differences to those retained, a factor that limits the generalizability of the results. However, tipping point analyses and the fact that there were no significant arm×attrition interactions on outcomes suggest that no sampling bias was introduced as a result of the attrition. Attrition could be a consequence of allowing users to be relatively anonymous and having a fast and accessible way of signing up for the study, lowering the threshold for engagement. The high attrition also means that the power to detect effect sizes was smaller than planned. In the between-group comparison of the self-help and therapist-guided arms at the 3-month follow-up, observed sample sizes would have given 80% power to detect an effect size of $d>0.29$, which is to be considered a small difference. However, we cannot rule out that the true difference is smaller than this. Adherence to the program was relatively low, with only 30.4% (119/391) completers in the self-help ICBT and 39.9% (154/386) in the therapist-guided ICBT, which is consistent with previous studies on ICBT (Multimedia Appendix 1). Higher adherence might improve the effects of the internet-based program. Owing to the web-based setting, the participants did not go through a clinical diagnostic interview, and some participants may not have been diagnosed as having alcohol dependence had an interview been included in the study design.

Future Directions

There is still a need for more studies on multisession internet-based interventions for harmful alcohol use and alcohol dependence, including studies with long-term follow-ups. No differences between the groups were found in number of participants that reported low-risk drinking at follow-ups. Only some internet help-seekers might need ICBT and therapist guidance to change their drinking when they use internet-based interventions. Others who did not improve might have benefited from more intensive support. A model of support-on-demand or accelerated care could be tested in future studies on the internet. One important challenge for future studies is to improve follow-up rates and adherence to interventions without reducing the willingness to use the interventions. Increased demands on users to identify themselves or have contact with a professional might make people who wish to remain anonymous or feel ashamed or stigmatized more reluctant to seek support [81]. Treatment-seeking increases the rates of recovery from alcohol dependence [82], and internet-based interventions seem to be a possible way to reach individuals currently not seeking treatment [83]; however, it is still unclear whether internet-based interventions actually increase treatment-seeking. Research on other psychiatric disorders [84] and on internet-based alcohol interventions so far [17] suggest that therapist-guided internet treatment has effects comparable with those of face-to-face treatment; however, more studies are needed that directly compare these interventions, as a recent study by our group has done [85]. In sparsely populated countries such as Sweden, where some people have to travel far to visit a clinic in person, psychological treatment [86], medical management [87], and after care [88] could, in part, be handled with internet-based interventions. More studies are needed to understand how internet-based interventions can be used effectively to improve treatment for people with alcohol dependence.

Conclusions

In this study, a therapist-guided ICBT program was not found to be more effective than the same program as a self-help ICBT for reducing alcohol consumption or other alcohol-related outcomes. In the short run, therapist-guided ICBT seems to be more effective than information. Only some internet help-seekers might need a multisession program and therapist guidance to change their drinking when they use internet-based interventions.

Acknowledgments

The authors would like to thank the therapists at Alkoholhjälpen—Rosette Lind, Liselott Sahlberg, and Karin Romberg—who delivered guided treatment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Controlled trials of internet cognitive behavioral therapy for alcohol use.
[DOCX File , 24 KB - jmir_v23i11e29666_app1.docx]

Multimedia Appendix 2

Demographic and clinical characteristics at baseline.
[DOCX File , 19 KB - jmir_v23i11e29666_app2.docx]

Multimedia Appendix 3
Observed mean, SD, and difference between groups.
[DOCX File , 22 KB - jmir_v23i11e29666_app3.docx]

Multimedia Appendix 4
Tipping point analyses of secondary outcomes.
[DOCX File , 13 KB - jmir_v23i11e29666_app4.docx]

Multimedia Appendix 5
Mixed models of secondary outcomes.
[DOCX File , 18 KB - jmir_v23i11e29666_app5.docx]

Multimedia Appendix 6
Correlations between changes in outcomes at follow-ups.
[DOCX File , 2954 KB - jmir_v23i11e29666_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1411 KB - jmir_v23i11e29666_app7.pdf]

References

1. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the global burden of disease study 2016. Lancet 2018 Sep 22;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]
2. Rehm J, Baliunas D, Borges GL, Graham K, Irving H, Kehoe T, et al. The relation between different dimensions of alcohol consumption and burden of disease: an overview. Addiction 2010 May;105(5):817-843 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.02899.x] [Medline: 20331573]
3. Rehm J, Marmet S, Anderson P, Gual A, Kraus L, Nutt DJ, et al. Defining substance use disorders: do we really need more than heavy use? Alcohol Alcohol 2013;48(6):633-640. [doi: 10.1093/alcalc/agt127] [Medline: 23926213]
4. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization; 1992.
5. Global status report on alcohol and health 2018. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2021-09-16]
6. Andréasson S, Danielsson A, Hallgren M. Severity of alcohol dependence in the swedish adult population: association with consumption and social factors. Alcohol 2013 Feb;47(1):21-25. [doi: 10.1016/j.alcohol.2012.10.001] [Medline: 23084032]
7. Kaner EF, Beyer F, Dickinson HO, Pienaar E, Campbell F, Schlesinger C, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2007 Apr 18(2):CD004148. [doi: 10.1002/14651858.CD004148.pub3] [Medline: 17443541]
8. Magill M, Ray L, Kiluk B, Hoadley A, Bernstein M, Tonigan JS, et al. A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: treatment efficacy by contrast condition. J Consult Clin Psychol 2019 Dec;87(12):1093-1105 [FREE Full text] [doi: 10.1037/ccp0000447] [Medline: 31599606]
9. Maisel NC, Blodgett JC, Wilbourne PL, Humphreys K, Finney JW. Meta-analysis of naltrexone and acamprosate for treating alcohol use disorders: when are these medications most helpful? Addiction 2013 Feb;108(2):275-293 [FREE Full text] [doi: 10.1111/j.1360-0443.2012.04054.x] [Medline: 23075288]
10. Rehm J, Anderson P, Manthey J, Shield KD, Struzzo P, Wojnar M, et al. Alcohol use disorders in primary health care: what do we know and where do we go? Alcohol Alcohol 2016 Jul;51(4):422-427. [doi: 10.1093/alcalc/agv127] [Medline: 26574600]
11. Degenhardt L, Glantz M, Evans-Lacko S, Sadikova E, Sampson N, Thornicroft G, World Health Organization's World Mental Health Surveys collaborators. Estimating treatment coverage for people with substance use disorders: an analysis of data from the world mental health surveys. World Psychiatry 2017 Oct;16(3):299-307 [FREE Full text] [doi: 10.1002/wps.20457] [Medline: 28941090]
12. Wallhed FS, Bakshi A, Andréasson S. Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. Subst Use Misuse 2014 May;49(6):762-769. [doi: 10.3109/10826084.2014.891616] [Medline: 24601784]

13. Probst C, Manthey J, Martinez A, Rehm J. Alcohol use disorder severity and reported reasons not to seek treatment: a cross-sectional study in european primary care practices. Subst Abuse Treat Prev Policy 2015 Aug 12;10:32 [FREE Full text] [doi: 10.1186/s13011-015-0028-z] [Medline: 26264215]
14. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010 Dec 07;8:82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]
15. Davidsson P, Thoresson A. Svenskarna och internet 2017. Internetstiftelsen. 2017. URL: https://2017.svenskarnaochinternet.se/ [accessed 2021-09-16]
16. Kypri K, Sitharthan T, Cunningham JA, Kavanagh DJ, Dean JI. Innovative approaches to intervention for problem drinking. Curr Opin Psychiatry 2005 May;18(3):229-234. [doi: 10.1097/01.yco.0000165591.75681.ab] [Medline: 16639145]
17. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
18. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: an individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec 18;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
19. Dedert EA, McDuffie JR, Stein R, McNiel JM, Kosinski AS, Freiermuth CE, et al. Electronic interventions for alcohol misuse and alcohol use disorders: a systematic review. Ann Intern Med 2015 Aug 4;163(3):205-214. [doi: 10.7326/M15-0285] [Medline: 26237752]
20. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
21. Sundström C, Blankers M, Khadjesari Z. Computer-based interventions for problematic alcohol use: a review of systematic reviews. Int J Behav Med 2017 Oct;24(5):646-658 [FREE Full text] [doi: 10.1007/s12529-016-9601-8] [Medline: 27757844]
22. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Dec 16;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
23. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
24. Sundström C, Gajecki M, Johansson M, Blankers M, Sinadinovic K, Stenlund-Gens E, et al. Guided and unguided internet-based treatment for problematic alcohol use - a randomized controlled pilot trial. PLoS One 2016 Jul 06;11(7):e0157817 [FREE Full text] [doi: 10.1371/journal.pone.0157817] [Medline: 27383389]
25. Brief DJ, Rubin A, Keane TM, Enggasser JL, Roy M, Helmuth E, et al. Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: a randomized clinical trial. J Consult Clin Psychol 2013 Oct;81(5):890-900. [doi: 10.1037/a0033697] [Medline: 23875821]
26. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. PLoS One 2011 Mar 09;6(3):e14740 [FREE Full text] [doi: 10.1371/journal.pone.0014740] [Medline: 21408060]
27. Boß L, Lehr D, Schaub MP, Paz CR, Riper H, Berking M, et al. Efficacy of a web-based intervention with and without guidance for employees with risky drinking: results of a three-arm randomized controlled trial. Addiction 2018 Apr;113(4):635-646 [FREE Full text] [doi: 10.1111/add.14085] [Medline: 29105879]
28. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
29. Hester RK, Delaney HD, Campbell W. ModerateDrinking.Com and moderation management: outcomes of a randomized clinical trial with non-dependent problem drinkers. J Consult Clin Psychol 2011 Apr;79(2):215-224 [FREE Full text] [doi: 10.1037/a0022487] [Medline: 21319896]
30. Brendryen H, Lund IO, Johansen AB, Riksheim M, Nesvåg S, Duckert F. Balance--a pragmatic randomized controlled trial of an online intensive self-help alcohol intervention. Addiction 2014 Feb;109(2):218-226. [doi: 10.1111/add.12383] [Medline: 24134709]
31. Cunningham JA. Comparison of two internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012 Aug 01;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]
32. Cunningham JA, Shorter GW, Murphy M, Kushnir V, Rehm J, Hendershot CS. Randomized controlled trial of a brief versus extended internet intervention for problem drinkers. Int J Behav Med 2017 Oct;24(5):760-767. [doi: 10.1007/s12529-016-9604-5] [Medline: 27770293]
33. Hester RK, Lenberg KL, Campbell W, Delaney HD. Overcoming addictions, a web-based application, and SMART recovery, an online and in-person mutual help group for problem drinkers, part 1: three-month outcomes of a randomized controlled trial. J Med Internet Res 2013 Jul 11;15(7):e134 [FREE Full text] [doi: 10.2196/jmir.2565] [Medline: 23846588]

34. Sinadinovic K, Wennberg P, Johansson M, Berman AH. Targeting individuals with problematic alcohol use via Web-based cognitive-behavioral self-help modules, personalized screening feedback or assessment only: a randomized controlled trial. Eur Addict Res 2014;20(6):305-318 [FREE Full text] [doi: 10.1159/000362406] [Medline: 25300885]
35. Sundström C, Eék N, Kraepelien M, Fahlke C, Gajecki M, Jakobson M, et al. High- versus low-intensity internet interventions for alcohol use disorders: results of a three-armed randomized controlled superiority trial. Addiction 2020 May;115(5):863-874 [FREE Full text] [doi: 10.1111/add.14871] [Medline: 31691413]
36. Lindström D, Sundberg-Petersson I, Adami J, Tönnesen H. Disappointment and drop-out rate after being allocated to control group in a smoking cessation trial. Contemp Clin Trials 2010 Jan;31(1):22-26. [doi: 10.1016/j.cct.2009.09.003] [Medline: 19758579]
37. Alkoholhjälpen. URL: https://alkoholhjalpen.se/ [accessed 2021-09-16]
38. Saunders J, Aasland O, Babor T, 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]
39. Johansson M, Sinadinovic K, Hammarberg A, Sundström C, Hermansson U, Andreasson S, et al. Web-based self-help for problematic alcohol use: a large naturalistic study. Int J Behav Med 2017 Oct;24(5):749-759 [FREE Full text] [doi: 10.1007/s12529-016-9618-z] [Medline: 27900733]
40. Finn S, Hammarberg A, Andreasson S. Treatment for alcohol dependence in primary care compared to outpatient specialist treatment-a randomized controlled trial. Alcohol Alcohol 2018 Jul 01;53(4):376-385. [doi: 10.1093/alcalc/agx126] [Medline: 29346473]
41. Miller W. Motivational Enhancement Therapy Manual: Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals with Alcohol Abuse and Dependence. Brattleboro, US: Echo Point Books and Media; 2014.
42. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. New York: Guilford Publications; 2012.
43. Monti P, Abrams DB, Kadden RM, Cooney NL. Treating Alcohol Dependence: A Coping Skills Training Guide. New York: Guilford Press; 2002.
44. Marlatt G, Donovan D. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. New York: Guilford press; 2005.
45. Sanchez-Craig M. Saying When: How to Quit Drinking Or Cut Down. Canada: Centre for Addiction and Mental Health; 2015.
46. Hester RK, Delaney HD. Behavioral self-control program for windows: results of a controlled clinical trial. J Consult Clin Psychol 1997 Aug;65(4):686-693. [doi: 10.1037//0022-006x.65.4.686] [Medline: 9256570]
47. Ahacic K, Nederfeldt L, Helgason AR. The national alcohol helpline in sweden: an evaluation of its first year. Subst Abuse Treat Prev Policy 2014 Jul 11;9:28 [FREE Full text] [doi: 10.1186/1747-597X-9-28] [Medline: 25015403]
48. Sobell L, Sobell M. Timeline follow-back. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption. Totowa, NJ: Humana Press; 1992.
49. Rueger SY, Trela CJ, Palmeri M, King AC. Self-administered web-based timeline followback procedure for drinking and smoking behaviors in young adults. J Stud Alcohol Drugs 2012 Sep;73(5):829-833 [FREE Full text] [doi: 10.15288/jsad.2012.73.829] [Medline: 22846247]
50. Thomas BA, McCambridge J. Comparative psychometric study of a range of hazardous drinking measures administered online in a youth population. Drug Alcohol Depend 2008 Jul 01;96(1-2):121-127. [doi: 10.1016/j.drugalcdep.2008.02.010] [Medline: 18406079]
51. Bergman H, Källmén H. Alcohol use among swedes and a psychometric evaluation of the alcohol use disorders identification test. Alcohol Alcohol 2002;37(3):245-251. [doi: 10.1093/alcalc/37.3.245] [Medline: 12003912]
52. Sinadinovic K, Wennberg P, Berman AH. Population screening of risky alcohol and drug use via internet and interactive voice response (IVR): a feasibility and psychometric study in a random sample. Drug Alcohol Depend 2011 Mar 01;114(1):55-60. [doi: 10.1016/j.drugalcdep.2010.09.004] [Medline: 21030164]
53. Vitesnikova J, Dinh M, Leonard E, Boufous S, Conigrave K. Use of AUDIT-C as a tool to identify hazardous alcohol consumption in admitted trauma patients. Injury 2014 Sep;45(9):1440-1444. [doi: 10.1016/j.injury.2014.01.004] [Medline: 24629701]
54. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington, DC: American Psychiatric Association; 2013.
55. van Hout B, Janssen M, Feng Y, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]
56. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]

57. Svanborg P, Asberg M. A comparison between the beck depression inventory (BDI) and the self-rating version of the montgomery asberg depression rating scale (MADRS). J Affect Disord 2001 May;64(2-3):203-216. [doi: 10.1016/s0165-0327(00)00242-1] [Medline: 11313087]
58. Holländare F, Andersson G, Engström I. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. J Med Internet Res 2010 Dec 19;12(5):e49 [FREE Full text] [doi: 10.2196/jmir.1392] [Medline: 21169165]
59. 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]
60. Dear BF, Titov N, Sunderland M, McMillan D, Anderson T, Lorian C, et al. Psychometric comparison of the generalized anxiety disorder scale-7 and the penn state worry questionnaire for measuring response during treatment of generalised anxiety disorder. Cogn Behav Ther 2011;40(3):216-227. [doi: 10.1080/16506073.2011.582138] [Medline: 21770844]
61. Bertholet N, Cheng DM, Palfai TP, Samet JH, Saitz R. Does readiness to change predict subsequent alcohol consumption in medical inpatients with unhealthy alcohol use? Addict Behav 2009 Aug;34(8):636-640 [FREE Full text] [doi: 10.1016/j.addbeh.2009.03.034] [Medline: 19428189]
62. Campbell A, Hemsley S. Outcome rating scale and session rating scale in psychological practice: clinical utility of ultra-brief measures. Clin Psychol 2009 Mar;13(1):1-9. [doi: 10.1080/13284200802676391]
63. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
64. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011 Mar 02;13(1):e26 [FREE Full text] [doi: 10.2196/jmir.1523] [Medline: 21371988]
65. Permutt T. Sensitivity analysis for missing data in regulatory submissions. Stat Med 2016 Jul 30;35(17):2876-2879. [doi: 10.1002/sim.6753] [Medline: 26567763]
66. Mehrotra DV, Liu F, Permutt T. Missing data in clinical trials: control-based mean imputation and sensitivity analysis. Pharm Stat 2017 Sep;16(5):378-392. [doi: 10.1002/pst.1817] [Medline: 28631331]
67. 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]
68. Enders CK. Analyzing longitudinal data with missing values. Rehabil Psychol 2011 Nov;56(4):267-288. [doi: 10.1037/a0025579] [Medline: 21967118]
69. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychol Addict Behav 2013 Mar;27(1):166-177 [FREE Full text] [doi: 10.1037/a0029508] [Medline: 22905895]
70. Horton NJ, Kim E, Saitz R. A cautionary note regarding count models of alcohol consumption in randomized controlled trials. BMC Med Res Methodol 2007 Feb 15;7:9 [FREE Full text] [doi: 10.1186/1471-2288-7-9] [Medline: 17302984]
71. Tasca GA, Gallop R. Multilevel modeling of longitudinal data for psychotherapy researchers: I. the basics. Psychother Res 2009 Jul;19(4-5):429-437. [doi: 10.1080/10503300802641444] [Medline: 19235088]
72. Sundström C, Kraepelien M, Eék N, Fahlke C, Kaldo V, Berman A. High-intensity therapist-guided internet-based cognitive behavior therapy for alcohol use disorder: a pilot study. BMC Psychiatry 2017 May 26;17(1):197 [FREE Full text] [doi: 10.1186/s12888-017-1355-6] [Medline: 28549424]
73. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. BMC Med Res Methodol 2013 Dec 06;13:150 [FREE Full text] [doi: 10.1186/1471-2288-13-150] [Medline: 24314204]
74. McCambridge J, Kypri K, McElduff P. Regression to the mean and alcohol consumption: a cohort study exploring implications for the interpretation of change in control groups in brief intervention trials. Drug Alcohol Depend 2014 Feb 01;135:156-159 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.11.017] [Medline: 24342421]
75. Clifford P, Maisto S, Davis C. Alcohol treatment research assessment exposure subject reactivity effects: part I. alcohol use and related consequences. J Stud Alcohol Drugs 2007 Jul;68(4):519-528. [doi: 10.15288/jsad.2007.68.519] [Medline: 17568955]
76. McCambridge J, Kypri K. Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials. PLoS One 2011;6(10):e23748 [FREE Full text] [doi: 10.1371/journal.pone.0023748] [Medline: 21998626]
77. Sinclair JM, Chambers SE, Manson CC. Internet support for dealing with problematic alcohol use: a survey of the soberistas online community. Alcohol Alcohol 2017 Mar 09;52(2):220-226 [FREE Full text] [doi: 10.1093/alcalc/agw078] [Medline: 28182197]
78. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]

79. Rehm J, Allamani A, Elekes Z, Jakubczyk A, Manthey J, Probst C, et al. Alcohol dependence and treatment utilization in europe - a representative cross-sectional study in primary care. BMC Fam Pract 2015 Jul 29;16:90 [FREE Full text] [doi: 10.1186/s12875-015-0308-8] [Medline: 26219430]
80. Storbjörk J, Room R. The two worlds of alcohol problems: who is in treatment and who is not? Addict Res Theory 2009 Jul 11;16(1):67-84. [doi: 10.1080/16066350701578136]
81. Ekström V, Johansson M. Choosing internet-based treatment for problematic alcohol use-why, when and how? Users' experiences of treatment online. Addict Sci Clin Pract 2020 Jun 29;15(1):22 [FREE Full text] [doi: 10.1186/s13722-020-00196-5] [Medline: 32600456]
82. Trim R, Schuckit M, Smith T. Predictors of initial and sustained remission from alcohol use disorders: findings from the 30-year follow-up of the san diego prospective study. Alcohol Clin Exp Res 2013 Aug;37(8):1424-1431 [FREE Full text] [doi: 10.1111/acer.12107] [Medline: 23458300]
83. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010 Dec 19;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
84. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? a systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
85. Johansson M, Sinadinovic K, Gajecki M, Lindner P, Berman AH, Hermansson U, et al. Internet-based therapy versus face-to-face therapy for alcohol use disorder, a randomized controlled non-inferiority trial. Addiction 2021 May;116(5):1088-1100 [FREE Full text] [doi: 10.1111/add.15270] [Medline: 32969541]
86. Kay-Lambkin FJ, Baker AL, Kelly BJ, Lewin TJ. It's worth a try: the treatment experiences of rural and urban participants in a randomized controlled trial of computerized psychological treatment for comorbid depression and alcohol/other drug use. J Dual Diagn 2012 Nov 08;8(4):262-276. [doi: 10.1080/15504263.2012.723315]
87. Devine E, Ryan M, Falk D, Fertig J, Litten R. An exploratory evaluation of take control: a novel computer-delivered behavioral platform for placebo-controlled pharmacotherapy trials for alcohol use disorder. Contemp Clin Trials 2016 Sep;50:178-185 [FREE Full text] [doi: 10.1016/j.cct.2016.08.006] [Medline: 27521807]
88. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

Abbreviations

AUDIT: alcohol use disorder identification test
CBT: cognitive behavioral therapy
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
EQ-5D-5L: EuroQol-5 dimensions
GAD-7: Generalized Anxiety Disorder Assessment-7 items
ICBT: internet-based cognitive behavioral therapy
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
MADRS-S: Montgomery Asberg Depression Rating Scale–Self-rated
MI: motivational interviewing
TLFB: timeline follow-back

Edited by R Kukafka; submitted 16.04.21; peer-reviewed by S Victor, N Nazari, R Kessler; comments to author 29.05.21; revised version received 23.06.21; accepted 27.07.21; published 24.11.21.

Please cite as:
Johansson M, Berman AH, Sinadinovic K, Lindner P, Hermansson U, Andréasson S
Effects of Internet-Based Cognitive Behavioral Therapy for Harmful Alcohol Use and Alcohol Dependence as Self-help or With Therapist Guidance: Three-Armed Randomized Trial
J Med Internet Res 2021;23(11):e29666
URL: https://www.jmir.org/2021/11/e29666
doi:10.2196/29666
PMID:34821563

Original Paper

Electrophysiological Brain Changes Associated With Cognitive Improvement in a Pediatric Attention Deficit Hyperactivity Disorder Digital Artificial Intelligence-Driven Intervention: Randomized Controlled Trial

Rafael Medina[1], MSc; Jaime Bouhaben[1], MSc; Ignacio de Ramón[1,2,3], MSc; Pablo Cuesta[2,4], PhD; Luis Antón-Toro[2,5], MSc; Javier Pacios[2,5], PhD; Javier Quintero[6], MD, PhD; Josep Antoni Ramos-Quiroga[7,8,9], MD; Fernando Maestú[2,5], PhD

[1]Sincrolab Ltd, Madrid, Spain

[2]Laboratory of Computational and Cognitive Neuroscience, Centre for Biomedical Technology, Polytechnic University of Madrid, Pozuelo de Alarcón, Spain

[3]Faculty of Health, Camilo Jose Cela University, Villafranca del Castillo, Spain

[4]Radiology Rehabilitation and Physiotherapy, Complutense University of Madrid, Madrid, Spain

[5]Department of Experimental Psychology, Faculty of Psychology, Complutense University of Madrid, Madrid, Spain

[6]Department of Psychiatry, University Hospital Infanta Leonor, Madrid, Spain

[7]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[8]Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute, Barcelona, Spain

[9]Biomedical Network Research Centre on Mental Health, Barcelona, Spain

Corresponding Author:
Ignacio de Ramón, MSc
Sincrolab Ltd
Prensa 7
Madrid, 28033
Spain
Phone: 34 630 364 425
Email: nacho@sincrolab.es

Abstract

Background: Cognitive stimulation therapy appears to show promising results in the rehabilitation of impaired cognitive processes in attention deficit hyperactivity disorder.

Objective: Encouraged by this evidence and the ever-increasing use of technology and artificial intelligence for therapeutic purposes, we examined whether cognitive stimulation therapy implemented on a mobile device and controlled by an artificial intelligence engine can be effective in the neurocognitive rehabilitation of these patients.

Methods: In this randomized study, 29 child participants (25 males) underwent training with a smart, digital, cognitive stimulation program (KAD_SCL_01) or with 3 commercial video games for 12 weeks, 3 days a week, 15 minutes a day. Participants completed a neuropsychological assessment and a preintervention and postintervention magnetoencephalography study in a resting state with their eyes closed. In addition, information on clinical symptoms was collected from the child´s legal guardians.

Results: In line with our main hypothesis, we found evidence that smart, digital, cognitive treatment results in improvements in inhibitory control performance. Improvements were also found in visuospatial working memory performance and in the cognitive flexibility, working memory, and behavior and general executive functioning behavioral clinical indexes in this group of participants. Finally, the improvements found in inhibitory control were related to increases in alpha-band power in all participants in the posterior regions, including 2 default mode network regions of the interest: the bilateral precuneus and the bilateral posterior cingulate cortex. However, only the participants who underwent cognitive stimulation intervention (KAD_SCL_01) showed a significant increase in this relationship.

Conclusions: The results seem to indicate that smart, digital treatment can be effective in the inhibitory control and visuospatial working memory rehabilitation in patients with attention deficit hyperactivity disorder. Furthermore, the relation of the inhibitory

control with alpha-band power changes could mean that these changes are a product of plasticity mechanisms or changes in the neuromodulatory dynamics.

Trial Registration: ISRCTN Registry ISRCTN71041318; https://www.isrctn.com/ISRCTN71041318

(J Med Internet Res 2021;23(11):e25466) doi:10.2196/25466

KEYWORDS

ADHD; cognitive stimulation; magnetoencephalography; artificial intelligence; Conners continuous performance test; KAD_SCL_01; AI; cognitive impairment; attention deficit hyperactivity disorder; pediatrics; children; rehabilitation

Introduction

Inhibitory control deficit is one of the core impairments in attention deficit hyperactivity disorder (ADHD) [1,2]. This deficit is directly related to the levels of impulsiveness present in the symptoms of ADHD [1,3-5] and produces difficulties in the everyday activities of those afflicted [6] while adversely affecting academic performance [7]. According to the literature reviewed, other impairments can be found in ADHD including the performance of cognitive processes, such as working memory [8], sustained attention [9,10], alternating attention [11], and planning [12,13].

In ADHD, the indices of inhibition, task switching, and emotional control appear to be related to relative power values of the alpha frequency band (7-13 Hz) in midline brain regions measured at resting state [14,15] and with performance in attentional tasks [16]. These patients consistently present a decrease in the alpha band in the central and posterior regions [17-24], as well as an increase in the theta frequency band (3-7 Hz) and the theta: beta ratio [17-21,25-28]. The decrease of the alpha band in regions that engage the default mode network (DMN; active network at resting state which includes the caudate nucleus, medial prefrontal cortex, posterior cingulate cortex, hippocampus, inferior parietal lobe, cerebellum, and precuneus) could modulate impairments in the functional connectivity of this network [29-33]. These impairments in the DMN also seem to be related to inhibitory control deficits [34,35].

Cognitive stimulation therapy appears to be effective in patients with ADHD [36]. The progressive increase of the workload in cognitive stimulation tasks is one of the main treatment dynamics of this type of therapy [37], and there are many examples of its effectiveness in ADHD and other disorders [38-42]. Its effectiveness seems to stem from the fact that these increases in the workload in cognitive tasks trigger an increase in long-distance connections supported by alpha and beta bands, and a decrease in short-distance connections supported by delta and theta bands [43-46].

Although the increase of the workload in cognitive stimulation tasks has shown promise in neurocognitive rehabilitation in children with ADHD, a case-based reasoning (CBR) system [47] that allows the adaptive workload to increase for each patient has never been used. The CRB system has been successful in various clinical areas [48-50], but its efficacy in a digital treatment for rehabilitation of neurocognitive alterations and its relation to electrophysiological dynamics and its efficacy in the rehabilitation of clinical alterations in ADHD remain unknown. With the aim of providing evidence, we examined whether a CRB digital training regimen would be effective in an ADHD child population after 12 weeks using the continuous performance test (CPT) inhibitory control measure as the main outcome. We hypothesized that after the intervention, the inhibitory control, as a core symptom of ADHD, would show a better performance and that this would be related to changes in the alpha band in the posterior regions and the DMN according to magnetoencephalography (MEG). We also tested whether treatment-produced changes in secondary outcomes would be related to ADHD and, finally, whether it could decrease the clinical symptoms associated with ADHD and change those behaviors related to executive functioning.

Methods

The study was approved by the local ethics committee of the San Carlos Hospital (Madrid, Spain). All legal representatives of the participants gave their written informed consent to participate in the study. This clinical trial is registered in the ISRCTN registry (ISRCTN71041318).

Participants

A total of 41 children diagnosed with combined-type ADHD (ADHD-C) were recruited (34 males). Contact with participants' legal guardians was made through health facilities, schools, and associations in the community of Madrid. Research staff first contacted those private and public clinical centers asking for permission and agreement to recruit. The order in which centers were contacted was at random. The following recruitment actions were performed: emailing study information, phone calls, and teleconferences and webinars with legal guardians summarizing study information. Participants' legal guardians who agreed to participate authorized communications with research staff. Eligibility criteria were checked by phone and email with legal guardians prior to visit 1. Before any other study activity, legal guardians read and signed an informed consent. There were no artificial intelligence (AI) requirements for the eligibility.

To be eligible, participants had to meet the following 5 criteria: (1) aged 8 to 11 years; (2) diagnosis of ADHD-C by an authorized professional (chartered psychiatrists at the medical college); (3) cessation of ADHD medication 3 days before each visit day, as, according to the technical specification of the drug methylphenidate (Concerta), it has a half-life of 3.5 hours (90% is excreted in urine and 1 to 35 in feces as a metabolite at 48-96 hours); (4) maintenance of the same level of medication during the at-home intervention period; and (5) compliance with the intervention protocol.

ADHD diagnosis was performed by accredited expert professionals following the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) criteria. These diagnostic criteria were the same across all participants. The average time from ADHD diagnosis confirmation to study enrollment was 2.58 (SD 1.21) years.

Participants meeting any of the following 5 exclusion criteria were dropped from the trial: (1) the initiation or abandonment of behavioral therapies or psychoactive drugs during the at-home intervention period; (2) motor difficulties which made the use of the mobile device (tablet or smartphone) impossible; (3) use of psychoactive drugs (such as benzodiazepines) which could have acted as a confounding factor, presence or suspicion of substance abuse for the past 6 months; (4) presence of blindness or uncorrected visual acuity difficulties; and (5) any additional psychological diagnosis.

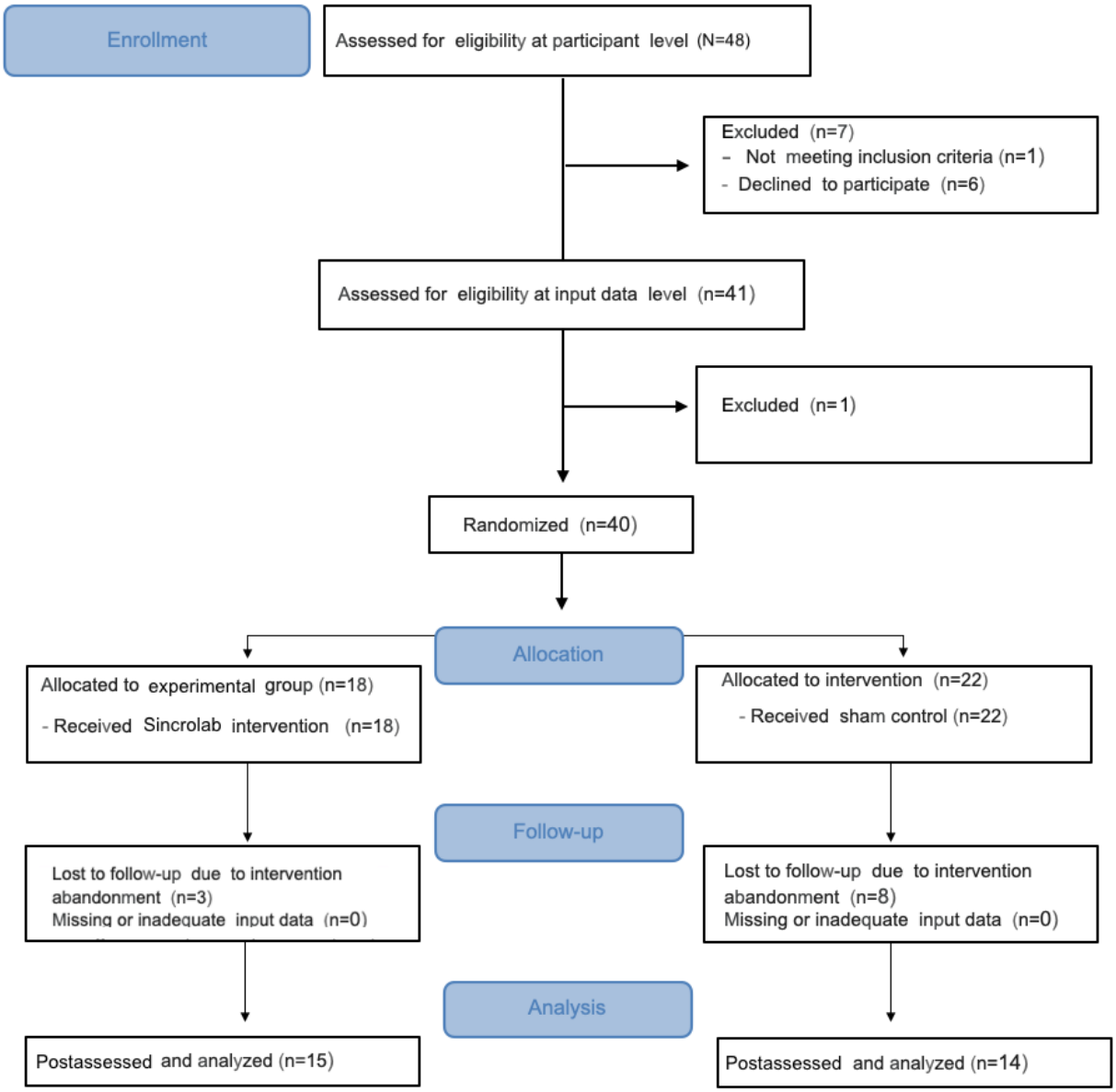
The inclusion criterion at the level of input data for the AI was a diagnosis of ADHD-C by an authorized professional in order to register the patient on the platform.

The use of other psychoactive drugs different from those approved by the Spanish Agency of Medicines and Medical Devices or European Medicines Agency for pediatric ADHD intervention (dextroamphetamine, levoamphetamine, lisdexamphetamine, methylphenidate, atomoxetine) also made participants ineligible for the study. The compliance was checked at the beginning and the end of the participant's participation through the child's legal guardians.

From the initial pool of 41 volunteers, 40 were randomly allocated into 1 of the 2 trial conditions (experimental or control). Of these, 28% (n=11) dropped out during the intervention period (control=8, experimental=3). One participant did not the meet inclusion criterion of stopping ADHD medication prior to treatment. The Consolidated Standards of Reporting Trials (CONSORT) 2010 flow diagram is presented in Figure 1.

From the final sample of 29 participants, 20 were taking pharmacological interventions (experimental=9, control=11), 7 participants were taking nonpharmacological interventions such as psychological interventions (experimental= 4, control=3), and 4 were taking both (experimental=2, control=2). The pharmacological interventions were based on methylphenidate (Concerta; n=7), methylphenidate (Equasym; n=7), methylphenidate hydrochloride (Medikinet; n=2), lisdexamphetamine (Elvanse; n=2), and methylphenidate hydrochloride (Rubifen; n=2).

Figure 1. Consolidated Standards of Reporting Trials flow diagram.



Experimental Design

This was single-center, parallel, single-blind, randomized controlled trial that examined a pediatric population (8-11 years) diagnosed with ADHD of combined presentation. It was conceptualized as a proof-of-concept study intended to assess the preliminary efficacy of a digital, videogame-like, cognitive stimulation therapy, as well as its safety and engagement. Proof-of-concept trials are useful in the framework of novel drugs and devices, so knowledge regarding their administration (eg, dosing, user instructions) may be acquired in small samples in order to develop larger clinical trials [51,52].

Digital Intervention

Experimental Condition

KAD_SCL_01 games are designed to work on different cognitive processes with an increase of the cognitive load following evidence that the brain's reconfiguration networks seem to be fixed by this type of training routine [43-46,53]. The 14 games which compose the KAD_SCL_01 cognitive intervention are described in Multimedia Appendix 1. The game level is adapted based on a case-based reasoning algorithm. This algorithm and the human-AI interaction are described in Multimedia Appendix 2.

Control Condition

Participants received a sham intervention composed of 3 videogames which were not specifically designed to improve

cognitive performance [54]. The specifications are described in Multimedia Appendix 3. The sham intervention tasks are accessible through Kongregate open-access platform (Kongragate Inc).

Main Outcome Measure and Magnetoencephalography

Main Outcome Measure

The main outcome measure of this study was the change in score found in the commission score from Conners CPT (CPT-III) between both groups' differences (pre- and postintervention) [55]. Commissions in CPT-III as a measure of inhibitory control was chosen as main outcome measure due to its use as an efficacy intervention measure in several previous studies about the methylphenidate effect in ADHD [56].

Magnetoencephalography

Neurophysiological data were acquired using a whole-head Elekta-Neuromag MEG system with 306 channels (Elekta AB) at the Center for Biomedical Technology (Madrid, Spain). MEG data were collected at a sampling frequency of 1000 Hz and online band-pass filtered between 0.1 Hz and 330 Hz.

Head shape was defined relative to 3 anatomical locations (nasion and bilateral preauricular points) using a 3D digitizer (Fastrak), and head motion was tracked through 4 head-position indicator coils attached to the scalp. Eye movements were monitored by a vertical electrooculogram assembly composed of a pair of bipolar electrodes.

Other Cognitive Outcome and Clinical Outcome Measures

The secondary cognitive outcome, aimed at measuring other several aspects of cognitive processing, and clinical questionnaires on ADHD behavioral symptoms and executive functioning in daily activities are included in Multimedia Appendix 4.

Safety and Compliance

Intervention safety was assessed through adverse events. Potential adverse events were monitored and recorded during the intervention period. Intervention dropouts were also recorded in order to assess compliance with intervention protocol.

Study Procedure

The study procedure occurred in 4 stages: recruitment and screening, preintervention assessment (visit 1), at-home intervention, and postintervention assessment (visit 2). Recruitment and screening were carried out as described in the Participants section. The details of the Al are described in Multimedia Appendix 5.

Preintervention and postintervention assessments were performed at the Center for Biomedical Technology, at the Technical University of Madrid. Assessments were carried out by a blinded Sincrolab researcher (JB) who only knew the number associated with the participant. Assessments including neuropsychological batteries and MEG recordings were administered in the same order as reported here. Questionnaires for the clinical outcome measures were filled out by the legal guardians. The cognitive assessment lasted for approximately 60 minutes. The resting-state MEG was also recorded during visits 1 and 2. The order in which participants received both was counterbalanced.

The intervention allocation was created by a nonblinded Sincrolab researcher (RM) and performed with a simple randomization function, with a ratio of 1:1 and an allocation probability of 0.50. Intervention allocation was performed once the eligibility criteria were met, according to the 2010 CONSORT statement [57].

The intervention was scheduled for 12 weeks, with 3 sessions (15-20 minutes each) per week in both groups. The whole intervention period was telematically monitored. Both the KAD_SCL_01 and sham control platforms allowed for a daily checking of performed sessions for a nonblinded Sincrolab researcher (RM). The number of weekly intervention sessions performed by the participants was monitored to ensure compliance with the 12-week intervention protocol. Safety and adequacy (the number of games played and the consecutive extreme punctuations of 0 or 100 in the performance, which could reflect an issue in the calibration of the AI outputs) were also assessed. Legal guardians were contacted by study staff in order to report any adverse event.

Right after the at-home intervention period was over, participants who achieved at least 80% completion of intervention sessions (28 alongside the 12 prescribed weeks) were appointed for postintervention assessment with same characteristics as the preintervention one. After the postintervention assessment, the participants who were allocated in the control arm were offered training with the KAD_SCL_01 for 12 weeks.

Statistical Analyses

Data analysis in this proof-of-concept randomized trial followed a per-protocol approach [58]. A per-protocol population was defined as any participant who had been randomly allocated to 1 of the 2 conditions (experimental or control), complied with at least an 80% completion of scheduled sessions (28 of 36), and had received the postintervention assessment.

Statistical Analyses of Cognitive Outcome Measures

Descriptive statistics of average, distribution shape, and scatter were calculated. Standardized statistics of asymmetry and kurtosis were used to assess the normality assumptions of each distribution. These standardized statistics are calculated by dividing the statistic between its SE.

Next, cognitive outcome measures which did not deviate from normality were adjusted to mixed-effects models. Each model was adjusted with a random intercept and fixed slope (due to the number of repeated measures). An unstructured covariance matrix (Sigma) was estimated for the random effect factor. Robust restricted maximum likelihood was chosen as the estimation method of preference due to its robustness with small samples and its capability to estimate an unbiased parameter matrix in the presence of missing values. A stepwise method was used for age as a demographic covariable in the main outcome's mixed model as a method applied to explicative models.

As the commission score from CPT-III was set as main outcome measure, only 1 comparison was performed (1 dependent variable). Therefore, no correction for multiplicity was applied. Regarding the rest of the cognitive outcome measures, every *P* value under significance α value of .05 was taken as statistically significant due to the exploratory nature of this pilot study. Still, *P* values were corrected for multiple comparisons under a Bonferroni correction within a statistical family. The outcome measures from the different cognitive processes (ie, visuospatial working memory) were treated as independent statistical families for Bonferroni adjustments.

Effect sizes greater than 0.4 (considered as the minimum practical effect size [59] in the experimental condition but not in the control condition) were highlighted. Likewise, for the main outcome, the predictive positive value (PPV) was estimated, as the small sample size could have led to overestimation of the effect size. Due to the novelty of this type of training methodology, a priori effect size and unspecified prestudy odds (*R*=0.5) were used in order to estimate the PPV.

Respondent analysis was also performed over the main cognitive outcome measure (commission score on CPT-III) in order to study the proportion of participants per intervention arm who achieved a pre-post difference of at least 0.64 SD, according to other literature [56]. Moreover, with consideration to this a priori effect and because the estimated sample size could not be achieved, post hoc power analysis for the mixed model's interaction component was carried out with 200 simulations, and PPV was computed following the procedure in Button et al [60].

Statistical Analyses of Clinical Outcome Measures

Clinical outcome measures were standardized according to Behavior Rating Inventory of Executive Function (BRIEF) and Evaluación del Trastorno por Deficit de Atención e Hiperactividad (EDAH) standardized scores (*T* scores). Paired-samples *t* tests were performed over each outcome measure and in each intervention group. Respondent analysis was also performed on the EDAH outcome measures by counting the proportion of participants who reached the cutoff point of pathology set by the interpretation of the EDAH manual for each condition. BRIEF and EDAH were treated as independent statistical families for Bonferroni adjustments.

Magnetoencephalography Signal Preprocessing and Statistical Analyses

With the intention of facilitating this paper's interpretation, signal preprocessing analyses are detailed in Multimedia Appendix 6.

Regarding the statistical analyses of the MEG preprocessed signal, the aim of this study was the detection of any robust correlation between power ratio values derived from the clusters of nodes localized in certain brain regions and CPT's commission ratio (CPT commission postintervention or CPT commission preintervention). The goal of this methodology included the extraction of any neurophysiological markers whose dynamic could be associated with the evolution of the inhibition-control performance. Such analysis relied on network-based statistics [61,62]. Clusters were built according to a criterion of spatial and frequency adjacency. Each cluster consisted of several adjacent nodes, which systematically showed a significant partial correlation (with age as the covariate) in at least 4 consecutive frequency steps (a 1-Hz interval) between their corresponding power ratio values and CPT ratio (Spearman correlation coefficient *P* value <.05). Importantly, all nodes within a cluster needed to show the same sign of the correlation coefficient for the cluster to be considered a functional unit. Only clusters involving at least 1% of the nodes (ie, a minimum of 12 nodes) in each frequency step were considered. Cluster-mass statistics were assessed through the sum of the Spearman ρ values across all nodes and significant frequency steps.

To control for multiple comparisons, the entire analysis pipeline was then repeated 5000 times, with the correspondence between power ratio estimates and CPT ratio being shuffled across participants. At each repetition, the maximum statistic of the surrogate clusters (in absolute value) was kept, creating a maximal null distribution that would ensure control of the familywise error rate at the cluster level. Cluster-mass statistics on each cluster in the original data set were compared with the same measure in the randomized data. The network-based statistics *P* value represented the proportion of the permutation distribution with cluster-mass statistic values greater or equal to the cluster-mass statistic value of the original data.

Power ratio values were averaged across all nodes and frequencies that belonged to the cluster. This average was considered to be the representative MEG marker value for that cluster and further participated in subsequent correlation analyses. Therefore, the statistics presented in the results section was derived from the correlation between the averaged power ratio value of each significant cluster and the corresponding CPT ratio for each participant. As already mentioned, correlations were first performed within the whole sample. In a second step, correlations between the average power ratio and the CPT commission ratio scores were performed independently for both intervention conditions within the sample (experimental and control). Statistical analyses were carried out using MATLAB R2020b (Mathworks Inc).

Sample Size Justification

A priori sample size was estimated to detect a standardized mean difference of 0.64 SD in the commission score from the CPT-III [56], with a significance level of α=.05 and a power of 0.8 (1-β=.8). The calculation procedure followed the sample size estimation for a 2-tailed, 2-samples mean difference with a correction factor for repeated measures [63]. The total sample size required was 56, but the actual sample was 29. Nevertheless, sample sizes of between 10 and 15 participants per condition are well-supported in similar literature [64-66].

The last enrolled participant ended study procedures in February 2020. With the COVID-19 crisis and the consequences in Spain (since March 2020), the study sponsor and principal investigator (FM) decided to stop the recruitment procedures due to difficulties and in order to assure protocol compliance in 2020. Therefore, assuming the exploratory nature of this pilot randomized trial, it was decided that the statistical analysis plan be applied to the presented sample.

Results

Demographic and Baseline Characteristics

Baseline demographics and other characteristics in each group, as well as the between-group comparison, are shown in Table 1. No significant differences were found between groups.

Table 1. Demographic characteristics in the experimental and control conditions.

Characteristic	Experimental group, (%)[a] (N=15)	Control group, (%)[a] (N=14)	*t* or X^2	*P* value
Age (years)	9.2 (1.21)	9.71 (1.33)	1.09	.27[b]
Males	13 (44.8)	12 (41.4)	0.005	.94[c]
Using medication	9 (31)	11 (37.9)	1.17	.28[c]
Receiving psychological treatment	4 (13.8)	3 (10,3)	0.11	.74[c]

[a]The characteristic of age is expressed as mean (SD).

[b]*P* values are from a *t* test (between-participant, 2-tailed).

[c]*P* values are from a chi-squared test (2-tailed).

Safety and Compliance

Three adverse events were reported by legal guardians during the at-home intervention period (Multimedia Appendix 7). Dropout (n=11) details are shown in Multimedia Appendix 8.

Main Outcome

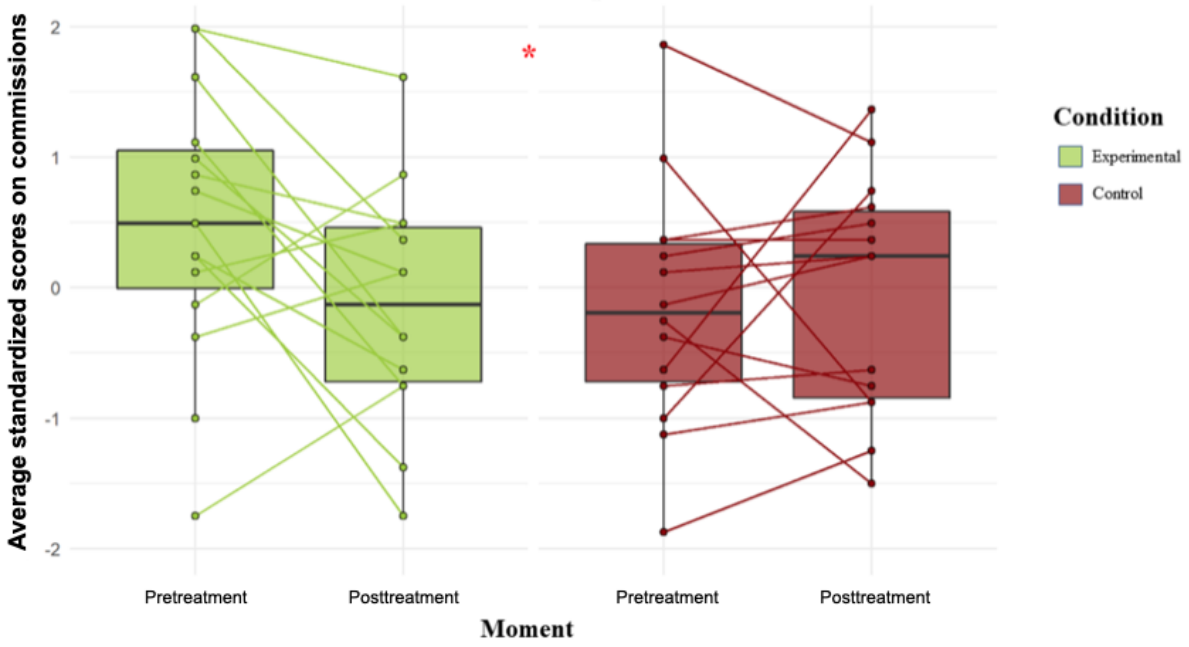
Descriptive statistics for the main outcome measure in each condition at each study period are shown in Table 2. The evolution trend (pre-post training) of each participant and the distribution of each group is shown in Figure 2. No statistically significant difference was found between the conditions (experimental and control) in the preintervention (baseline) measures (t_{27}=1.72; *P*=.10). Critical ratios for skewness indicated no deviations in skewness or kurtosis in the normal distribution.

Table 2. Descriptive statistics for main outcome measure commission score on Conners continuous performance test (CPT-III).

Descriptive statistic	Treatment group		Control group	
	Pretreatment	Posttreatment	Pretreatment	Posttreatment
Mean (SD)	53.87 (8.37)	47.80 (8.21)	48.79 (7.53)	49.64 (7.32)
Asymmetry	–0.37	–0.17	0.28	–0.09
Kurtosis	–0.61	–0.76	–0.33	–1.52
CR[a] asymmetry	–0.32	–0.15	0.23	–0.08
CR kurtosis	–0.27	–0.34	–0.14	–0.66

[a]CR: critical ratio.

Figure 2. Main efficacy outcome: individual and average change in commission errors from Conners continuous performance test per condition.



Mixed-effects models for main outcome measured with and without interaction effects were estimated with the robust restricted maximum likelihood procedure. The stepwise introduction of the condition-period interaction effect significantly improved the model adjustment (X^2_1=4.596; *P*=.03). Standardized mean difference (β estimator) for the condition-period interaction effect in the final model (Figure 2) was statistically different from 0 (β=.86; SE 0.39; t_{27}=2.21; *P*=.04). The standardized mean differences (β estimators) for model comparison (baseline model to final model) are shown in Multimedia Appendix 9. Comparison criteria (Akaike information criterion and Bayesian information criterion) between the models, in addition to model performance statistics (R^2 and adjusted R^2), are also reported in Multimedia Appendix 9. The graphical diagnosis for the final model with interaction effect is shown in Multimedia Appendix 10.

Pre-post standardized mean differences per condition were calculated as Hedges g statistic for effect size. A large pre-post standardized mean difference (g>|0.4|) was found in the experimental group (g=–0.62), but not in the control group (g=0.1). A high PPV (PPV=0.81) was found to be related with the pre-post standardized mean difference.

Respondent analysis for the main outcome measure shows that 53.33% (8/15) of the experimental participants (KAD_SCL_01 intervention) achieved the a priori clinically meaningful effect: an improvement of at least 0.64 standardized points. In the control arm, this percentage was just 21.42% (3/14). More details about respondent analysis are shown in Multimedia Appendix 11.

A post hoc power analysis yielded a statistical power of 43% (1 – β=0.43) for the detection of the condition-period interaction effect and a PPV of 0.81. A priori effect size was used to simulate post hoc power rather than observed effect size.

Magnetoencephalography Outcomes

A significant cluster (*P*=.04) was found in the frequency interval (11.67-13.33 Hz) mainly comprising the posterior regions of the brain (see Figure 3A and Table 3).

The power ratio in all frequencies of this interval negatively correlated with the CPT ratio across the whole sample (ρ=–0.562; *P*=.003). The maximum cluster size was found at 12-12.33 Hz (51 nodes). The cluster size oscillated between a minimum of 50 nodes at the beginning of the frequency range and 16 at the end of that frequency range (see Figure 3B). Furthermore, 12 Hz showed the highest average correlation coefficient value across all nodes of the cluster ρ= –0.547).

The correlation between the CPT commission ratio and the power ratio (11.67-13.33 Hz) in the interval within the cluster generated in the previous step remained significant for the experimental group (ρ=–0.783; *P*=.004; Figure 3C) but not for the control group (ρ=–0.358; *P*=.21; Figure 3C).

Figure 3. Brain region whose magnetoencephalography alpha power (11.67-13.33 Hz) was found significantly correlated with CPT commission ratio. (A) Brain regions within the significant cluster (depicted in blue). (B) Evolution of the cluster size through the different frequency steps (maximum size at 11.75 Hz). (C) Scatter plot showing the Spearman correlation coefficient between the cluster's average power ratio and CPT commission ratio and each subgroup of the sample. CPT: continuous performance test; Freqs: frequency steps.

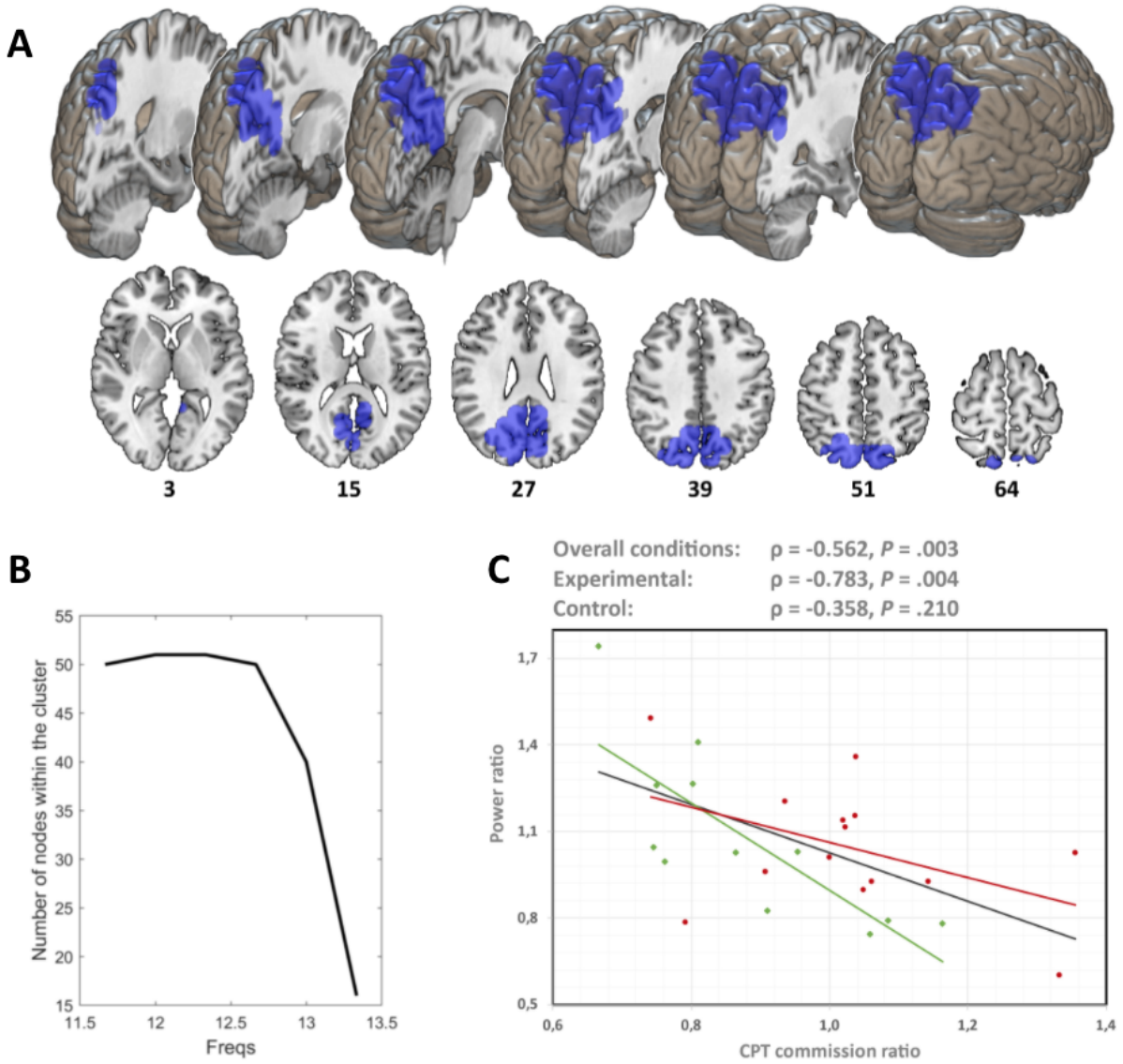


Table 3. The automated anatomical labeling atlas ROIs[a] that were partially captured by the significant cluster.[b]

ROI	Portion of ROI occupied, n/N (%)[c]
Left precuneus	11/28 (39.29)
Right precuneus	8/21 (38.10)
Left cuneus	7/11 63.64)
Right cuneus	7/13 (53.85)
Right superior parietal gyrus	6/18 (33.33)
Left cingulate gyrus, posterior part	3/5 (60.00)
Right superior occipital lobe	3/10 (30.00)
Left superior parietal gyrus	2/16 (12.50)
Right cingulate gyrus, posterior part	1/4 (25.00)
Left calcarine fissure and surrounding cortex	1/20 (5.00)
Left superior occipital lobe	1/11 (9.09)
Right middle occipital lobe	1/17 (5.88)

[a]ROI: region of interest.

[b]Regions of interest are from the Anatomical Labeling Atlas that are part of the significant cluster where the continuous performance test commission ratio correlates with power in the alpha band.

[c]N is the number of magnetoencephalography sources in our head model that are contained within the ROI volume; n indicates how many sources, among the corresponding N, are enclosed within the significant cluster; and % is the percentage of each ROI that was captured by that cluster.

Other Cognitive and Clinical Outcomes

Descriptive analysis for each secondary outcome measure is shown in Supplementary Material (Multimedia Appendix 12). No statistically significant differences were found between conditions (experimental and control) in preintervention measurement.

The mixed-effects model analysis was performed the for main outcome measure. Only, the backward span score (from the Corsi block-tapping test) as a dependent variable (X^2_1=4.64; *P*=.03) was significant. The standardized mean difference for the condition–moment interaction effect (β estimator) in the final model was statistically different from 0 (β =–.84; SE 0.38, t_{27}=–2.24; *P*=.03). Multimedia Appendix 13 shows the graphical representation of the average pre- and postintervention standardized scores for each intervention group (experimental and control) in this outcome measure. Standardized mean differences and CIs for cognitive secondary outcomes are shown in Multimedia Appendix 14, as classified by the cognitive process each measures (inhibitory control, cognitive flexibility, working memory, short-term memory, attention, speed processing, and verbal fluency).

Effect sizes of g>0.4 in the experimental group but not the control group were found in 19 cognitive secondary outcome measures, plus the main outcome. In contrast, only 1 cognitive secondary outcome measure showed a greater effect size (g>0.4) in the control group compared to the experimental one. See Multimedia Appendix 15 for the complete analysis.

The results in the parent version of the BRIEF questionnaire showed statistically significant pre-post mean differences, favoring the KAD_SCL_01 intervention participants in shifting score (t_{14}=2.32; *P*=.03), working memory score (t_{14}= 2.43, *P*=.02), behavioral composite index (t_{14}=2.62, *P*=.02), and general executive composite index (t_{14}= 2.7, *P*=.01). No significant differences in the sham intervention group were found. The experimental arm (KAD_SCL_01 intervention) showed statistically significant pre-post mean differences in all EDAH measures (hyperactivity score *P*=.05), inattention score (*P*=.001), behavior disorder score (*P*=.001), and global score (*P*=.001). The control arm also showed statistically significant pre-post mean differences in inattention score (*P*=.001), hyperactivity + inattention, and global score (*P*=.002), but not in hyperactivity or behavior disorder score. Respondent analysis, descriptive analysis, *t* statistics, *P* values, CIs, and respondent percentage per score in the EDAH scale are detailed in Multimedia Appendix 16.

No statistically significant differences were found between conditions (experimental and control) in the preintervention measurement either in the BRIEF or the EDAH outcome measures.

Discussion

Empirical evidence points suggests that cognitive stimulation based on progressive workload increments leads to improvements in cognitive performance [38-42,67], along with beneficial regulation of cortical activity patterns [43-46,53].

The results in our study indicate that cognitive intervention triggers significant improvements in inhibitory control in child and adolescent patients with ADHD as measured by Conners CPT-III. Moreover, this improvement in inhibitory control seems to be similar to that found in pharmacological studies on the effectiveness of methylphenidate [68]. Meanwhile, the effect size of our study (g=0.62) is consistent with that found in the meta-analysis by Losier et al [56] on the effectiveness of the

drugs used in ADHD. Therefore the digital treatment proposed in the present study could be a therapeutic option complementary to the pharmacological route.

Despite there being no significant differences between the groups in the measure of previous treatment (P=.09), the possible differences between both groups could be producing a type I error or false-positive result. However, as observed in Figure 2, within the range of 1 to –1 SD, 7 of the 10 patients who received the KAD_SCL_01 treatment show an improvement in their performance (70%). On the contrary, 6 of 9 participants belonging to the control group, in the same range, show worse scores in the postintervention measure.

Although several studies have reported that digital cognitive exercises do not show effects superior to those found in other commercial video games not intended for therapeutic uses [37,69,70], these findings, like those reported by Davis et al [41] and Kollins et al [42], seem to indicate that adaptive digital training, built on a proven empirical basis, could be effective for the treatment of ADHD.

The relationship between alpha-band power and performance in tasks involving attentional and inhibitory control processes has been published in recent publications [14-16,71,72]. In order to clarify the relationship between the changes in inhibitory control and the possible changes in alpha band power—given its association with performance in inhibitory control tasks [14,15]—we completed a MEG registry of the participants of both groups. The results seem to indicate that there is a direct relationship (ρ=–0.56; P=.003) between the improvement in inhibitory control and the alpha-band power in the posterior brain areas. These changes in the power of brain oscillations appear to be associated with brain plasticity processes [73]**,** as well as changes in the dynamics of neuromodulators such as dopamine [74] that are affected in these patients. This relationship remained significant when the experimental group (ρ= –0.78; P=.004) was analyzed separately, but this was not the case with the control group (ρ= –0.35; P=.21). This suggests that the improvements produced in the experimental group are strongly associated with the previously mentioned plasticity and neuromodulation phenomena.

Although this was intended as a power study, we believe our results are relevant to the functional connectivity literature due to the participation of the precuneus and posterior cingulate cortex in the main cluster examined of this paper. These 2 regions of interest conform to the posterior part of the DMN. Consequently, the alpha-power increment linked with the CPT decrement may be associated, under our interpretation, with an improvement in the functional integrity of the DMN. The decrease in alpha-band power in regions of the DMN could be mediating the impairments present in ADHD in the functional connectivity of this network [75-77] and its neurocognitive correlate [34,35,78]. This effect seems to be due to frequency band having special importance in the communication between the regions of this network [33]. In this regard, the data from Sonuga-Barke and Castellanos [79] seems to indicate that the decrease in connectivity between the regions of the DMN generates interference in the task-oriented network, producing impairments in the performance of patients with ADHD [79]. From this perspective, this digital cognitive stimulation intervention, based on progressive workload increases governed by CBR algorithms, might be effective for the treatment of ADHD.

One of the secondary objectives of the study was to measure the effectiveness of the intervention on a set of cognitive processes that are usually part of the ADHD deficits. The results indicate that cognitive intervention triggers improvements in visuospatial working memory total score. Moreover, practical minimum effect size (g> 0.4) was observed in visuospatial working memory span and in visuospatial working memory total score, while numeric working memory total score or numeric working memory span seemed to be not be affected by the treatment. These results are similar to those found by another digital study, which based its training on empirical principles [41]. In our study, no significant differences were found in other measures. However, these results might be due to a type II or false-negative error, since small sample sizes frequently generate this type of error [80]. Indeed, we found a practical minimum effect size in 12 of the 42 secondary outcome indices, which is a possible indicator of treatment efficacy. In contrast, the control group showed equivalent effects in just 1of the 42 secondary measures (see Multimedia Appendix 15 for detailed information).

Finally, in this study, the parent version of the BRIEF questionnaire was used to obtain a measure of executive functions in everyday life. The results indicate that cognitive training triggers significant changes in flexibility, working memory, and the composite indices of both behavior and executive functions. These changes appear to be similar to the effects of methylphenidate-based pharmacological treatments [81] and treatments administered by clinical professionals for executive functions [82].

In the ADHD rating scale, in the overall rating, 60% (9/15) of the participants who underwent cognitive training exceeded the cutoff point (<30), compared to 21% (3/14) who worked with commercial video games, which seems to indicate that this type of cognitive training may have positive effects on the behavioral impact of the disorder.

In conclusion, this study reports the preliminary results of a digital cognitive stimulation intervention in a population with ADHD. The results suggest that such treatment is effective at improving inhibitory control and visuospatial working memory in patients with ADHD. Moreover, this improvement was observed in the executive measures of daily life and was associated with a reduction of symptoms.

The main limitation of the study relates to the small size of the sample (N=29) compared to the a priori calculated sample size (N=56). Consequently, the statistical power was lower than the one desired a priori. Therefore, these results must be interpreted as the first evidence of a digital treatment using CBR algorithms, and more extensive studies are needed to confirm the findings of this proof-of-concept study.

Acknowledgments

This study was funded by Sincrolab and partly funded by the Centre for the Development of Industrial Technology of the Spanish Ministry of Economy, Industry, and Competitiveness.

Authors' Contributions

Author contribution: RM, JB, IR, JQ, and FM designed research; JB, LAT, and PC conducted the research; RM, JB, LAT, and PC analyzed data; and RM, JB, IR, LAT, JQ, JARQ, and FM wrote the paper.

Conflicts of Interest

Sincrolab provided financial support in the form of salaries used for partial salary support for the authors. PC received punctual financial support for carrying out the magnetoencephalography analysis. Sincrolab participated in the study design, data analysis, decision to publish, and preparation of the manuscript. IR is the cofounder of Sincrolab. JQ and JARQ are members of the Scientific Board of Sincrolab. JQ is also a shareholder of Instituto Neuroconductual de Madrid Ltd and a speaker on the advisory board for Takeda & Jansen. He also receives investigation funding from the Carlos III Health Institute. JARQ was on the speakers' bureau and/or acted as a consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, Bristol Myers Squibb, Medice, Rubió, Uriach, and Raffo in the last 3 years. He also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial, and Medice. The Department of Psychiatry chaired by JARQ received unrestricted educational and research support from the following companies in the last 3 years: Janssen-Cilag, Shire, Oryzon, Roche, Psious, and Rubió. RM and JB are employees of Sincrolab. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Dynamics, integrated cognitive processes, and hierarchical or multilevel structure of Sincrolab's games.
[PDF File (Adobe PDF File), 1011 KB - jmir_v23i11e25466_app1.pdf]

Multimedia Appendix 2
Case-based reasoning explanation and human-artificial intelligence interaction.
[PDF File (Adobe PDF File), 310 KB - jmir_v23i11e25466_app2.pdf]

Multimedia Appendix 3
Description of sham control intervention.
[PDF File (Adobe PDF File), 84 KB - jmir_v23i11e25466_app3.pdf]

Multimedia Appendix 4
Cognitive and clinical secondary outcome measures.
[PDF File (Adobe PDF File), 137 KB - jmir_v23i11e25466_app4.pdf]

Multimedia Appendix 5
Study procedure diagram.
[PDF File (Adobe PDF File), 71 KB - jmir_v23i11e25466_app5.pdf]

Multimedia Appendix 6
Magnetoencephalography signal preprocessing procedure.
[PDF File (Adobe PDF File), 101 KB - jmir_v23i11e25466_app6.pdf]

Multimedia Appendix 7
Adverse events.
[PDF File (Adobe PDF File), 29 KB - jmir_v23i11e25466_app7.pdf]

Multimedia Appendix 8
Dropouts' details.
[PDF File (Adobe PDF File), 28 KB - jmir_v23i11e25466_app8.pdf]

Multimedia Appendix 9
Comparison between performance and standardized mean differences (beta estimators) for each main outcome model.

[PDF File (Adobe PDF File), 36 KB - jmir_v23i11e25466_app9.pdf]

Multimedia Appendix 10
Graphical diagnosis for main outcome's final model with interaction effect.
[PDF File (Adobe PDF File), 27 KB - jmir_v23i11e25466_app10.pdf]

Multimedia Appendix 11
Comparison of individual performance and clinical effect.
[PDF File (Adobe PDF File), 10 KB - jmir_v23i11e25466_app11.pdf]

Multimedia Appendix 12
Descriptive statistics for secondary outcome measures.
[PDF File (Adobe PDF File), 61 KB - jmir_v23i11e25466_app12.pdf]

Multimedia Appendix 13
Efficacy outcome in visuospatial working memory: mean change in backward span score from Corsi block-tapping test. Significant differences were found for condition × moment interaction effect in this cognitive measure (β=–.84; SE=0.38; <italic>t</italic>₂₇=–2.24; <italic>P</italic>=.03).
[PDF File (Adobe PDF File), 44 KB - jmir_v23i11e25466_app13.pdf]

Multimedia Appendix 14
Standardized mean differences for interaction effects in secondary outcome measures.
[PDF File (Adobe PDF File), 77 KB - jmir_v23i11e25466_app14.pdf]

Multimedia Appendix 15
Pre-post standardized mean differences (Hedges g) per treatment group.
[PDF File (Adobe PDF File), 44 KB - jmir_v23i11e25466_app15.pdf]

Multimedia Appendix 16
Descriptives, <italic>t</italic> statistics, <italic>P</italic> values, CIs, and respondent percentage per score in the Evaluación del Trastorno por Deficit de Atención e Hiperactividad scale.
[PDF File (Adobe PDF File), 21 KB - jmir_v23i11e25466_app16.pdf]

Multimedia Appendix 17
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 631 KB - jmir_v23i11e25466_app17.pdf]

References

1. 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 Jul 01;57(11):1336-1346. [doi: 10.1016/j.biopsych.2005.02.006] [Medline: 15950006]
2. Coghill DR, Seth S, Matthews K. A comprehensive assessment of memory, delay aversion, timing, inhibition, decision making and variability in attention deficit hyperactivity disorder: advancing beyond the three-pathway models. Psychol. Med 2013 Oct 31;44(9):1989-2001. [doi: 10.1017/s0033291713002547]
3. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [Medline: 9000892]
4. Brown TE. Executive functions and attention deficit hyperactivity disorder: implications of two conflicting views. International Journal of Disability, Development and Education 2007 Jan 22;53(1):35-46 [FREE Full text] [doi: 10.1080/10349120500510024]
5. Barkley RA. Is executive functioning deficient in ADHD? It depends on your definitions and measures. The ADHD Report 2011 Aug;19(4):1-9,16. [doi: 10.1521/adhd.2011.19.4.1]
6. Duff CT, Sulla EM. Measuring executive function in the differential diagnosis of attention-deficit/hyperactivity disorder: Does it really tell us anything? Appl Neuropsychol Child 2015 Sep 25;4(3):188-196 [FREE Full text] [doi: 10.1080/21622965.2013.848329] [Medline: 25257827]
7. Best JR, Miller PH, Naglieri JA. Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. Learn Individ Differ 2011 Aug;21(4):327-336 [FREE Full text] [doi: 10.1016/j.lindif.2011.01.007] [Medline: 21845021]

8. Fenollar-Cortés J, Navarro-Soria I, González-Gómez C, García-Sevilla J. Cognitive Profile for Children with ADHD by Using WISC-IV: Subtype Differences? // Detección de perfiles cognitivos mediante WISC-IV en niños diagnosticados de TDAH: ¿Existen diferencias entre subtipos? REV PSICODIDACT 2014 Nov 24;20(1):157-176 [FREE Full text] [doi: 10.1387/revpsicodidact.12531]
9. Byrne JM, Bawden HN, DeWolfe NA, Beattie TL. Clinical assessment of psychopharmacological treatment of preschoolers with ADHD. J Clin Exp Neuropsychol 1998 Oct;20(5):613-627. [doi: 10.1076/jcen.20.5.613.1121] [Medline: 10079039]
10. Zalsman G, Pumeranz O, Peretz G, Ben-Dor DH, Dekel S, Horesh N, et al. Attention patterns in children with attention deficit disorder with or without hyperactivity. ScientificWorldJournal 2003 Dec 13;3:1093-1107 [FREE Full text] [doi: 10.1100/tsw.2003.94] [Medline: 14625396]
11. Rodriguez-Jimenez R, Avila C, Jimenez-Arriero MA, Ponce G, Monasor R, Jimenez M, et al. Impulsivity and sustained attention in pathological gamblers: influence of childhood ADHD history. J Gambl Stud 2006 Dec;22(4):451-461. [doi: 10.1007/s10899-006-9028-2] [Medline: 16912931]
12. Seidman LJ, Biederman J, Monuteaux MC, Weber W, Faraone SV. Neuropsychological functioning in nonreferred siblings of children with attention deficit/hyperactivity disorder. Journal of Abnormal Psychology 2000 May;109(2):252-265 [FREE Full text] [doi: 10.1037/0021-843x.109.2.252]
13. Álvarez-Arboleda L, Rodríguez-Arocho W, Moreno-Torres M. Evaluación neurocognoscitiva del trastorno por déficit de atención e hiperactividad. IRD Éditions 2003:83-90 [FREE Full text] [doi: 10.4000/books.irdeditions.25484]
14. Swartwood JN, Swartwood MO, Lubar JF, Timmermann DL. EEG differences in ADHD-combined type during baseline and cognitive tasks. Pediatric Neurology 2003 Mar;28(3):199-204. [doi: 10.1016/s0887-8994(02)00514-3]
15. Zhang D, Johnstone SJ, Roodenrys S, Luo X, Li H, Wang E, et al. The role of resting-state EEG localized activation and central nervous system arousal in executive function performance in children with Attention-Deficit/Hyperactivity Disorder. Clin Neurophysiol 2018 Jun;129(6):1192-1200. [doi: 10.1016/j.clinph.2018.03.009] [Medline: 29653296]
16. Shi T, Li X, Song J, Zhao N, Sun C, Xia W, et al. EEG characteristics and visual cognitive function of children with attention deficit hyperactivity disorder (ADHD). Brain Dev 2012 Dec;34(10):806-811. [doi: 10.1016/j.braindev.2012.02.013] [Medline: 22459254]
17. Clarke AR, Barry RJ, Bond D, McCarthy R, Selikowitz M. Effects of stimulant medications on the EEG of children with attention-deficit/hyperactivity disorder. Psychopharmacology (Berl) 2002 Nov;164(3):277-284. [doi: 10.1007/s00213-002-1205-0] [Medline: 12424551]
18. Clarke A, Barry R, McCarthy R, Selikowitz M, Brown C. EEG evidence for a new conceptualisation of attention deficit hyperactivity disorder. Clinical neurophysiology 2002;113(7):1036-1044. [doi: 10.1016/s1388-2457(02)00115-3] [Medline: 12088697]
19. Clarke AR, Barry RJ, McCarthy R, Selikowitz M, Croft RJ. EEG differences between good and poor responders to methylphenidate in boys with the inattentive type of attention-deficit/hyperactivity disorder. Clinical Neurophysiology 2002 Aug;113(8):1191-1198. [doi: 10.1016/s1388-2457(02)00147-5]
20. Clarke AR, Barry RJ, McCarthy R, Selikowitz M, Brown CR, Croft RJ. Effects of stimulant medications on the EEG of children with Attention-Deficit/Hyperactivity Disorder Predominantly Inattentive type. International Journal of Psychophysiology 2003 Feb;47(2):129-137. [doi: 10.1016/s0167-8760(02)00119-8]
21. Bresnahan SM, Barry RJ, Clarke AR, Johnstone SJ. Quantitative EEG analysis in dexamphetamine-responsive adults with attention-deficit/hyperactivity disorder. Psychiatry Res 2006 Mar 28;141(2):151-159. [doi: 10.1016/j.psychres.2005.09.002] [Medline: 16343642]
22. Barry RJ, Clarke AR, Hajos M, McCarthy R, Selikowitz M, Bruggemann JM. Acute atomoxetine effects on the EEG of children with attention-deficit/hyperactivity disorder. Neuropharmacology 2009 Dec;57(7-8):702-707. [doi: 10.1016/j.neuropharm.2009.08.003] [Medline: 19698723]
23. Choi H, Park S, Kim KK, Lee K, Rhyu H. Acute effects of aerobic stretching, health and happiness improving movement exercise on cortical activity of children. J Exerc Rehabil 2016 Aug;12(4):320-327 [FREE Full text] [doi: 10.12965/jer.1632602.301] [Medline: 27656629]
24. Huang C, Huang C, Hung C, Tsai Y, Chang Y, Wu C, et al. Effects of Acute Exercise on Resting EEG in Children with Attention-Deficit/Hyperactivity Disorder. Child Psychiatry Hum Dev 2018 Dec;49(6):993-1002. [doi: 10.1007/s10578-018-0813-9] [Medline: 29872997]
25. Hermens DF, Williams LM, Clarke S, Kohn M, Cooper N, Gordon E. Responses to methylphenidate in adolescent AD/HD: evidence from concurrently recorded autonomic (EDA) and central (EEG and ERP) measures. Int J Psychophysiol 2005 Oct;58(1):21-33. [doi: 10.1016/j.ijpsycho.2005.03.006] [Medline: 15936104]
26. Kamida A, Shimabayashi K, Oguri M, Takamori T, Ueda N, Koyanagi Y, et al. EEG power spectrum analysis in children with ADHD. Yonago Acta Med 2016 Jul;59(2):169-173 [FREE Full text] [Medline: 27493489]
27. Markovska-Simoska S, Pop-Jordanova N. Quantitative EEG in children and adults with attention deficit hyperactivity disorder: comparison of absolute and relative power spectra and theta/beta ratio. Clin EEG Neurosci 2017 Jan;48(1):20-32. [doi: 10.1177/1550059416643824] [Medline: 27170672]
28. Clarke AR, Barry RJ, Johnstone SJ, McCarthy R, Selikowitz M. EEG development in attention deficit hyperactivity disorder: from child to adult. Clin Neurophysiol 2019 Aug;130(8):1256-1262. [doi: 10.1016/j.clinph.2019.05.001] [Medline: 31163371]

29. de Pasquale F, Della Penna S, Snyder AZ, Lewis C, Mantini D, Marzetti L, et al. Temporal dynamics of spontaneous MEG activity in brain networks. Proc Natl Acad Sci U S A 2010 Mar 30;107(13):6040-6045 [FREE Full text] [doi: 10.1073/pnas.0913863107] [Medline: 20304792]
30. Brookes MJ, Woolrich M, Luckhoo H, Price D, Hale JR, Stephenson MC, et al. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. Proc Natl Acad Sci U S A 2011 Oct 04;108(40):16783-16788 [FREE Full text] [doi: 10.1073/pnas.1112685108] [Medline: 21930901]
31. Tang W, Liu H, Douw L, Kramer MA, Eden UT, Hämäläinen MS, et al. Dynamic connectivity modulates local activity in the core regions of the default-mode network. Proc Natl Acad Sci USA 2017 Aug 21;114(36):9713-9718 [FREE Full text] [doi: 10.1073/pnas.1702027114]
32. Hacker CD, Snyder AZ, Pahwa M, Corbetta M, Leuthardt EC. Frequency-specific electrophysiologic correlates of resting state fMRI networks. Neuroimage 2017 Apr 01;149:446-457 [FREE Full text] [doi: 10.1016/j.neuroimage.2017.01.054] [Medline: 28159686]
33. Samogin J, Liu Q, Marino M, Wenderoth N, Mantini D. Shared and connection-specific intrinsic interactions in the default mode network. Neuroimage 2019 Oct 15;200:474-481 [FREE Full text] [doi: 10.1016/j.neuroimage.2019.07.007] [Medline: 31280013]
34. Greicius MD, Krasnow B, Reiss AL, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A 2003 Jan 07;100(1):253-258 [FREE Full text] [doi: 10.1073/pnas.0135058100] [Medline: 12506194]
35. Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex 2009 Jan;19(1):72-78 [FREE Full text] [doi: 10.1093/cercor/bhn059] [Medline: 18403396]
36. Catalá-López F, Hutton B, Núñez-Beltrán A, Page MJ, Ridao M, Macías Saint-Gerons D, et al. The pharmacological and non-pharmacological treatment of attention deficit hyperactivity disorder in children and adolescents: a systematic review with network meta-analyses of randomised trials. PLoS One 2017;12(7):e0180355 [FREE Full text] [doi: 10.1371/journal.pone.0180355] [Medline: 28700715]
37. Cortese S, Ferrin M, Brandeis D, Buitelaar J, Daley D, Dittmann RW, European ADHD Guidelines Group (EAGG). Cognitive training for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. J Am Acad Child Adolesc Psychiatry 2015 Mar;54(3):164-174 [FREE Full text] [doi: 10.1016/j.jaac.2014.12.010] [Medline: 25721181]
38. Tamm L, Epstein JN, Peugh JL, Nakonezny PA, Hughes CW. Preliminary data suggesting the efficacy of attention training for school-aged children with ADHD. Dev Cogn Neurosci 2013 May;4:16-28 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.004] [Medline: 23219490]
39. Tamm L, Nakonezny PA. Metacognitive executive function training for young children with ADHD: a proof-of-concept study. Atten Defic Hyperact Disord 2015 Oct;7(3):183-190 [FREE Full text] [doi: 10.1007/s12402-014-0162-x] [Medline: 25559877]
40. Spencer-Smith M, Klingberg T. Benefits of a working memory training program for inattention in daily life: a systematic review and meta-analysis. PLoS One 2016;11(11):e0167373 [FREE Full text] [doi: 10.1371/journal.pone.0167373] [Medline: 27875585]
41. Davis NO, Bower J, Kollins SH. Proof-of-concept study of an at-home, engaging, digital intervention for pediatric ADHD. PLoS One 2018;13(1):e0189749 [FREE Full text] [doi: 10.1371/journal.pone.0189749] [Medline: 29324745]
42. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RSE, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. The Lancet Digital Health 2020 Apr;2(4):e168-e178. [doi: 10.1016/s2589-7500(20)30017-0]
43. Kitzbichler MG, Henson RNA, Smith ML, Nathan PJ, Bullmore ET. Cognitive effort drives workspace configuration of human brain functional networks. Journal of Neuroscience 2011 Jun 01;31(22):8259-8270 [FREE Full text] [doi: 10.1523/jneurosci.0440-11.2011]
44. Fornito A, Harrison BJ, Zalesky A, Simons JS. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. Proc Natl Acad Sci U S A 2012 Jul 31;109(31):12788-12793 [FREE Full text] [doi: 10.1073/pnas.1204185109] [Medline: 22807481]
45. Arbabshirani MR, Havlicek M, Kiehl KA, Pearlson GD, Calhoun VD. Functional network connectivity during rest and task conditions: a comparative study. Hum Brain Mapp 2013 Dec;34(11):2959-2971 [FREE Full text] [doi: 10.1002/hbm.22118] [Medline: 22736522]
46. Finc K, Bonna K, Lewandowska M, Wolak T, Nikadon J, Dreszer J, et al. Transition of the functional brain network related to increasing cognitive demands. Hum Brain Mapp 2017 Jul;38(7):3659-3674 [FREE Full text] [doi: 10.1002/hbm.23621] [Medline: 28432773]
47. Custers EJFM, Durning SJ, ten Cate O. Principles and Practice of Case-based Clinical Reasoning Education. Berlin: Springer; 2018.
48. Marling CR, Petot GJ, Sterling LS. Integrating Case-Based and Rule-Based Reasoning to Meet Multiple Design Constraints. Computational Intell 1999 Aug;15(3):308-332 [FREE Full text] [doi: 10.1111/0824-7935.00095]

49. Marling C, Whitehouse P. Case-based reasoning in the care of alzheimer's disease patients. In: David W. Aha, Ian Watson, eds.Case-based reasoning research and development. ICCBR 2001. lecture notes in computer science. Berlin, Heidelberg: Springer; 2001.
50. Reddy M, Pesl P, Xenou M, Toumazou C, Johnston D, Georgiou P, et al. Clinical safety and feasibility of the advanced bolus calculator for type 1 diabetes based on case-based reasoning: A 6-week nonrandomized single-arm pilot study. Diabetes Technol Ther 2016 Aug;18(8):487-493. [doi: 10.1089/dia.2015.0413] [Medline: 27196358]
51. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
52. Kendig CE. What is proof of concept research and how does it generate epistemic and ethical categories for future scientific practice? Sci Eng Ethics 2016 Jun;22(3):735-753. [doi: 10.1007/s11948-015-9654-0] [Medline: 26009258]
53. Finc K, Bonna K, He X, Lydon-Staley DM, Kühn S, Duch W, et al. Author Correction: Dynamic reconfiguration of functional brain networks during working memory training. Nat Commun 2020 Jul 30;11(1):3891 [FREE Full text] [doi: 10.1038/s41467-020-17747-8] [Medline: 32732869]
54. Mishra J, Anguera JA, Gazzaley A. Video games for neuro-cognitive optimization. Neuron 2016 Apr 20;90(2):214-218 [FREE Full text] [doi: 10.1016/j.neuron.2016.04.010] [Medline: 27100194]
55. Conners K. Conners Continuous Performance Test (Conners CPT 3) & Conners Continuous Auditory Test of Attention (Conners CATA): Technical Manual. Toronto, Canada: Multi-Health Systems; 2014.
56. Losier BJ, McGrath PJ, Klein RM. Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: a meta-analytic review. J Child Psychol Psychiatry 1996 Dec;37(8):971-987. [doi: 10.1111/j.1469-7610.1996.tb01494.x] [Medline: 9119944]
57. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
58. Boutis K, Willan A. Intention-to-treat and per-protocol analysis. Canadian Medical Association Journal 2011 Apr 04;183(6):696 [FREE Full text] [doi: 10.1503/cmaj.111-2034]
59. Ferguson CJ. Is Psychological Research Really as Good as Medical Research? Effect Size Comparisons between Psychology and Medicine. Review of General Psychology 2009 Jun 01;13(2):130-136. [doi: 10.1037/a0015103]
60. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013 May;14(5):365-376. [doi: 10.1038/nrn3475] [Medline: 23571845]
61. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. Neuroimage 2010 Dec;53(4):1197-1207 [FREE Full text] [doi: 10.1016/j.neuroimage.2010.06.041] [Medline: 20600983]
62. Fornito A, Zalesky A, Bullmore E. Fundamentals of Brain Network Analysis. Amsterdam: Elsevier; 2016.
63. Twisk J. Applied Longitudinal Data Analysis for Epidemiology. New York: Cambridge University Press; 2013.
64. Castellanos NP, Paúl N, Ordóñez VE, Demuynck O, Bajo R, Campo P, et al. Reorganization of functional connectivity as a correlate of cognitive recovery in acquired brain injury. Brain 2010 Aug;133(Pt 8):2365-2381. [doi: 10.1093/brain/awq174] [Medline: 20826433]
65. Looi CY, Duta M, Brem A, Huber S, Nuerk H, Cohen Kadosh R. Combining brain stimulation and video game to promote long-term transfer of learning and cognitive enhancement. Sci Rep 2016 Mar 23;6:22003 [FREE Full text] [doi: 10.1038/srep22003] [Medline: 26902664]
66. Nyquist JB, Lappin JS, Zhang R, Tadin D. Perceptual training yields rapid improvements in visually impaired youth. Sci Rep 2016 Nov 30;6:37431 [FREE Full text] [doi: 10.1038/srep37431] [Medline: 27901026]
67. Soveri A, Antfolk J, Karlsson L, Salo B, Laine M. Working memory training revisited: A multi-level meta-analysis of n-back training studies. Psychon Bull Rev 2017 Aug;24(4):1077-1096. [doi: 10.3758/s13423-016-1217-0] [Medline: 28116702]
68. Wang L, Hsiao C, Huang Y, Chiang Y, Ree S, Chen Y, et al. Association of salivary dehydroepiandrosterone levels and symptoms in patients with attention deficit hyperactivity disorder during six months of treatment with methylphenidate. Psychoneuroendocrinology 2011 Oct;36(8):1209-1216. [doi: 10.1016/j.psyneuen.2011.02.014] [Medline: 21411231]
69. Lawlor-Savage L, Goghari VM. Dual N-back working memory training in healthy adults: a randomized comparison to processing speed training. PLoS One 2016;11(4):e0151817 [FREE Full text] [doi: 10.1371/journal.pone.0151817] [Medline: 27043141]
70. Melby-Lervåg M, Redick TS, Hulme C. Working memory training does not improve performance on measures of intelligence or other measures of "Far transfer":evidence from a meta-analytic review. Perspect Psychol Sci 2016 Jul;11(4):512-534 [FREE Full text] [doi: 10.1177/1745691616635612] [Medline: 27474138]
71. Lenartowicz A, Delorme A, Walshaw PD, Cho AL, Bilder RM, McGough JJ, et al. Elctroencephalography correlates of spatial working memory deficits in attention-deficit/hyperactivity disorder: vigilance, encoding, and maintenance. Journal of Neuroscience 2014 Jan 22;34(4):1171-1182 [FREE Full text] [doi: 10.1523/jneurosci.1765-13.2014]

72. Deiber M, Hasler R, Colin J, Dayer A, Aubry J, Baggio S, et al. Linking alpha oscillations, attention and inhibitory control in adult ADHD with EEG neurofeedback. Neuroimage Clin 2020;25:102145 [FREE Full text] [doi: 10.1016/j.nicl.2019.102145] [Medline: 31911342]
73. Nunez PL. Toward a quantitative description of large-scale neocortical dynamic function and EEG. Behav Brain Sci 2000 Jul;23(3):371-98; discussion 399. [doi: 10.1017/s0140525x00003253] [Medline: 11301576]
74. Takahashi S, Ohmiya M, Honda S, Ni K. The KCNH3 inhibitor ASP2905 shows potential in the treatment of attention deficit/hyperactivity disorder. PLoS One 2018;13(11):e0207750 [FREE Full text] [doi: 10.1371/journal.pone.0207750] [Medline: 30462746]
75. Tian L, Jiang T, Liang M, Zang Y, He Y, Sui M, et al. Enhanced resting-state brain activities in ADHD patients: a fMRI study. Brain Dev 2008 May;30(5):342-348. [doi: 10.1016/j.braindev.2007.10.005] [Medline: 18060712]
76. Sokunbi MO, Fung W, Sawlani V, Choppin S, Linden DE, Thome J. Resting state fMRI entropy probes complexity of brain activity in adults with ADHD. Psychiatry Res 2013 Dec 30;214(3):341-348. [doi: 10.1016/j.pscychresns.2013.10.001] [Medline: 24183857]
77. Zhao Q, Li H, Yu X, Huang F, Wang Y, Liu L, et al. Abnormal resting-state functional connectivity of insular subregions and disrupted correlation with working memory in adults with attention deficit/hyperactivity disorder. Front Psychiatry 2017 Oct 11;8:200 [FREE Full text] [doi: 10.3389/fpsyt.2017.00200] [Medline: 29075206]
78. Tomasi D, Volkow ND. Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. Biol Psychiatry 2012 Mar 01;71(5):443-450 [FREE Full text] [doi: 10.1016/j.biopsych.2011.11.003] [Medline: 22153589]
79. Sonuga-Barke EJS, Castellanos FX. Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. Neurosci Biobehav Rev 2007;31(7):977-986. [doi: 10.1016/j.neubiorev.2007.02.005] [Medline: 17445893]
80. Greenberg R, Daniels S, Flanders W, Eley J, Boring J. Interpretation of epidemiologic literature. In: Medical Epidemiology. New York: McGraw-Hill; 2001:175-188.
81. Wolraich ML, Greenhill LL, Pelham W, Swanson J, Wilens T, Palumbo D, et al. Randomized, controlled trial of oros methylphenidate once a day in children with attention-deficit/hyperactivity disorder. Pediatrics 2001 Oct;108(4):883-892. [doi: 10.1542/peds.108.4.883] [Medline: 11581440]
82. Shuai L, Daley D, Wang Y, Zhang J, Kong Y, Tan X, et al. Executive function training for children with attention deficit hyperactivity disorder. Chin Med J (Engl) 2017 Mar 05;130(5):549-558 [FREE Full text] [doi: 10.4103/0366-6999.200541] [Medline: 28229986]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ADHD-C: attention deficit hyperactivity disorder combined type
AI: artificial intelligence
BRIEF: Behavior Rating Inventory of Executive Function
CBR: case-based reasoning
CONSORT: Consolidated Standards of Reporting Trials
CPT: continuous performance test
CPT-III: Conner continuous performance test
DMN: default mode network
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EDAH: Evaluación del Trastorno por Deficit de Atención e Hiperactividad
MEG: magnetoencephalography
PPV: predictive positive value

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by A Lenartowicz, B Nievas Soriano; comments to author 18.11.20; revised version received 11.12.20; accepted 13.05.21; published 26.11.21.

Please cite as:
Medina R, Bouhaben J, de Ramón I, Cuesta P, Antón-Toro L, Pacios J, Quintero J, Ramos-Quiroga JA, Maestú F
Electrophysiological Brain Changes Associated With Cognitive Improvement in a Pediatric Attention Deficit Hyperactivity Disorder Digital Artificial Intelligence-Driven Intervention: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e25466
URL: https://www.jmir.org/2021/11/e25466
doi:10.2196/25466
PMID:34842533

Review

Mechanisms of Change in Digital Health Interventions for Mental Disorders in Youth: Systematic Review

Matthias Domhardt[1], MSc; Sophie Engler[1], BSc; Hannah Nowak[1], MSc; Arne Lutsch[1], BSc; Amit Baumel[2], PhD; Harald Baumeister[1], PhD

[1]Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

[2]Department of Community Mental Health, University of Haifa, Haifa, Israel

Corresponding Author:
Matthias Domhardt, MSc
Department of Clinical Psychology and Psychotherapy
Ulm University
Lise-Meitner-Str. 16
Ulm, D-89081
Germany
Phone: 49 73150 32803
Fax: 49 73150 32809
Email: matthias.domhardt@uni-ulm.de

Abstract

Background: Digital health interventions (DHIs) are efficacious for several mental disorders in youth; however, integrated, evidence-based knowledge about the mechanisms of change in these interventions is lacking.

Objective: This systematic review aims to comprehensively evaluate studies on mediators and mechanisms of change in different DHIs for common mental disorders in children and adolescents.

Methods: A systematic literature search of the electronic databases Cochrane Central Register of Controlled Trials, Embase, MEDLINE, and PsycINFO was conducted, complemented by backward and forward searches. Two independent reviewers selected studies for inclusion, extracted the data, and rated the methodological quality of eligible studies (ie, risk of bias and 8 quality criteria for process research).

Results: A total of 25 studies that have evaluated 39 potential mediators were included in this review. Cognitive mediators were the largest group of examined intervening variables, followed by a broad range of emotional and affective, interpersonal, parenting behavior, and other mediators. The mediator categories with the highest percentages of significant intervening variables were the groups of affective mediators (4/4, 100%) and combined cognitive mediators (13/19, 68%). Although more than three-quarters of the eligible studies met 5 or more quality criteria, causal conclusions have been widely precluded.

Conclusions: The findings of this review might guide the empirically informed advancement of DHIs, contributing to improved intervention outcomes, and the discussion of methodological recommendations for process research might facilitate mediation studies with more pertinent designs, allowing for conclusions with higher causal certainty in the future.

(J Med Internet Res 2021;23(11):e29742) doi:10.2196/29742

KEYWORDS

children and adolescents; mental disorders; mediator; mechanisms of change; digital health intervention; psychotherapy; mobile phone

Introduction

Background

Mental disorders in children and adolescents are common, with prevalence rates ranging from 10% to 20% worldwide [1]. These disorders contribute substantially to the global burden of disease in youth [2], and about half of all mental disorders across the life span have their onset in adolescence [3]. Hence, early psychotherapeutic interventions are essential to counteract the risk of chronification and prevent possible negative long-term effects [1]. However, a substantial proportion of children and adolescents with mental disorders do not receive adequate psychotherapeutic or psychosocial care [4-6] owing to different

individual and structural barriers to treatment uptake [7]. Furthermore, the availability of mental health care is often insufficient to adequately meet treatment demands, particularly in rural regions [8] and low-income countries [9].

Digital health interventions (DHIs), such as internet- and mobile-based interventions with a psychotherapeutic focus (DHI_{PSY}), offer the possibility of addressing some barriers to treatment uptake and might contribute to extending mental health care, given their various presumed advantages, such as possible cost- and time-efficient use, independence from spatial and temporal circumstances, potential anonymity, high degrees of flexibility, and autonomy for users. These assets may be especially important during the COVID-19 pandemic and allow for continued mental health care despite contact restrictions and physical distancing [10]. Furthermore, in view of the fact that youth are particular familiar with digital devices (so-called *digital natives*; Children in a digital world [11]), the use of DHIs might be especially appealing to this younger age group [12]. DHIs can be distinguished and characterized based on their theoretical basis, the type of technical implementation, the area of application, and the extent of accompanying human support [13,14]. The type and the intensity of *guidance* in DHIs can vary on the continuum from (1) pure self-help interventions without any human support (so-called *unguided interventions*) to (2) interventions with some support (*guided interventions*), to (3) videoconference-based psychotherapy with the internet as the sole communication medium between therapists and patients [15].

The efficacy of DHI_{PSY} for some common mental disorders in children and adolescents has been established using meta-analyses [12,16-19]. For example, Vigerland et al [19] evaluated the efficacy of internet-based cognitive behavioral therapy for a range of mental disorders, including anxiety, depression, behavioral problems, obsessive-compulsive disorder, and some somatic disorders such as chronic pain and insomnia. This meta-analysis revealed a moderate, aggregated effect size favoring internet-based cognitive behavioral therapy over waitlist (g=0.62, 95% CI 0.41-0.84; P<.001). In contrast, Hollis et al [20] appraised the evidence on the efficacy of DHI_{PSY} for other mental disorders, including attention-deficit/hyperactivity disorder, autism, psychotic disorders, and eating disorders in their review as uncertain and necessitating future research regarding moderators of intervention effects. Current empirical knowledge suggests that older children and adolescents benefit more from DHI_{PSY} than younger children [12,20]. In addition, the well-established finding that guided interventions are more efficacious than unguided interventions in adults [21,22] also seems to apply for DHI_{PSY} in youth [12,20]. Complementing the evidence base and representing another major area of application, DHIs with a focus on health promotion (DHI_{HP}), for example, on alcohol consumption or other lifestyle and health behaviors, revealed a considerably smaller effect size (Cohen d=0.14, 95% CI 0.00-0.27) [23] when compared with DHI_{PSY} with a genuine psychotherapeutic foundation such as cognitive behavioral therapy (g=0.72, 95% CI 0.55-0.90; P<.001) [16].

Given this rather heterogeneous body of research regarding the efficacy and effectiveness of DHIs, comprising various interventions with different theoretical orientations, foci, and delivery modes, for mental health issues in children and adolescents, it seems both timely and worthwhile to investigate the presumed working mechanisms in these technology-delivered interventions. This is because evidence-based knowledge on the mediators and mechanisms of change (specific for different approaches of DHI_{PSY} and DHI_{HP}) can inform intervention development and mental health care practices, illustrating pathways to more powerful intervention packages and improved outcomes [24-26]. The first step in understanding the underlying processes in DHIs is to analyze the mediators. A mediator is an intervening variable that can explain the statistical relationship between an independent variable (eg, a DHI) and a dependent variable (eg, a symptom change) [25], and can thereby potentially point to a mechanism through which an intervention achieves its effects. Although various methods for statistical mediation analysis are available (eg, MacKinnon et al [27]), comprising different approaches such as latent growth curve modeling [28] or structural equation modeling [29], the seminal approach of Baron and Kenny [30] is still one of the most applied procedures to evaluate the intervening variable effect of a potential mediator, despite having received criticism with regard to some limitations, such as low statistical power, difficulties in the assessment of multiple mediators, or quantification of the mediation effect magnitude [24]. Although statistical mediation may be established either with the so-called causal-steps approach by Baron and Kenny [30] or by more recent methods correcting some of its limitations (eg, Kraemer et al [31]), it is important to consider that mediators might be identical to a mechanism of change (ie, the actual process responsible for change), but might also be a proxy for 1 or more other variables that do not explain the hypothesized mechanism [25]. Thus, to determine the degree of validity that a mediator is actually representative of for being considered a true change mechanism, Kazdin proposed several quality criteria for psychotherapy process research [25] that can be consulted when assessing the scope and justification of causal inferences: (1) Strong association (among treatment, mediator, and outcome), (2) specificity (a mediator accounts for the indirect effect of treatment on outcome), (3) consistency (the association must be replicable), (4) experimental manipulation (use of either a randomized controlled trial [RCT] design where the intervention variable is manipulated or an experimental design where the mediator itself is directly manipulated), (5) timeline or temporality (the intervention must lead to changes in the mediator, which must temporally precede changes in the outcome), (6) gradient (ie, a dose-response relationship: greater activation of the mediator is associated with greater change in the outcome), and (7) plausibility or coherence (the proposed mediator must be embedded in a plausible theoretical framework).

The evidence base for the mediators and mechanisms of change in conventional face-to-face psychotherapies for children and adolescents is scarce, and only a few studies have been designed to investigate the therapeutic processes in these interventions

[32]. For example, 2 systematic reviews dedicated to evaluating the mechanisms of change found that only a small number of eligible primary studies actually conducted mediation analyses, with only 9% (6/67) [33] and 17% (8/46) of included clinical trials [34] attempting to evaluate mediation effects. In addition, Schmidt and Schimmelmann [35] reviewed the empirical literature on mediators in psychotherapeutic interventions for common mental disorders in youth and concluded that most eligible studies evaluated mediators referring to the parent-child interaction (eg, family cohesion and parental support), next to mediators within the patient (eg, self-efficacy, motivation, coping, interpersonal skills, as well as changes in dysfunctional cognitions and negative emotions) and characteristics of the intervention (eg, duration of treatment, number of patient-therapist contacts, and application of specific intervention techniques). The included studies revealed inconsistent patterns of mediation effects related to therapist factors, such as flexibility, adherence to treatment, or therapeutic alliance [35]. Moreover, as central conceptual and methodological requirements for mediation analyses were often not met by studies in this review, causal inferences were widely precluded, necessitating future process research with higher methodological rigor [35].

Objectives

Although research on DHIs is a fast growing field [36] and might also offer intriguing opportunities for psychotherapeutic process research [24,37], we are not aware of any systematic review of the mediators and mechanisms of change in DHI_{PSY} and DHI_{HP} for common mental disorders in youth published to this point. Therefore, this study aimed to: (1) systematically review the current state of research on mediators and mechanisms of change in various DHIs for mental disorders in children and adolescents, (2) identify mediators and potential mechanisms of change in these interventions, and (3) evaluate the methodological quality of eligible mediation studies according to the quality criteria for process research mentioned above.

Methods

This systematic review was reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [38] and was a priori registered with the Open Science Framework [39].

Eligibility Criteria

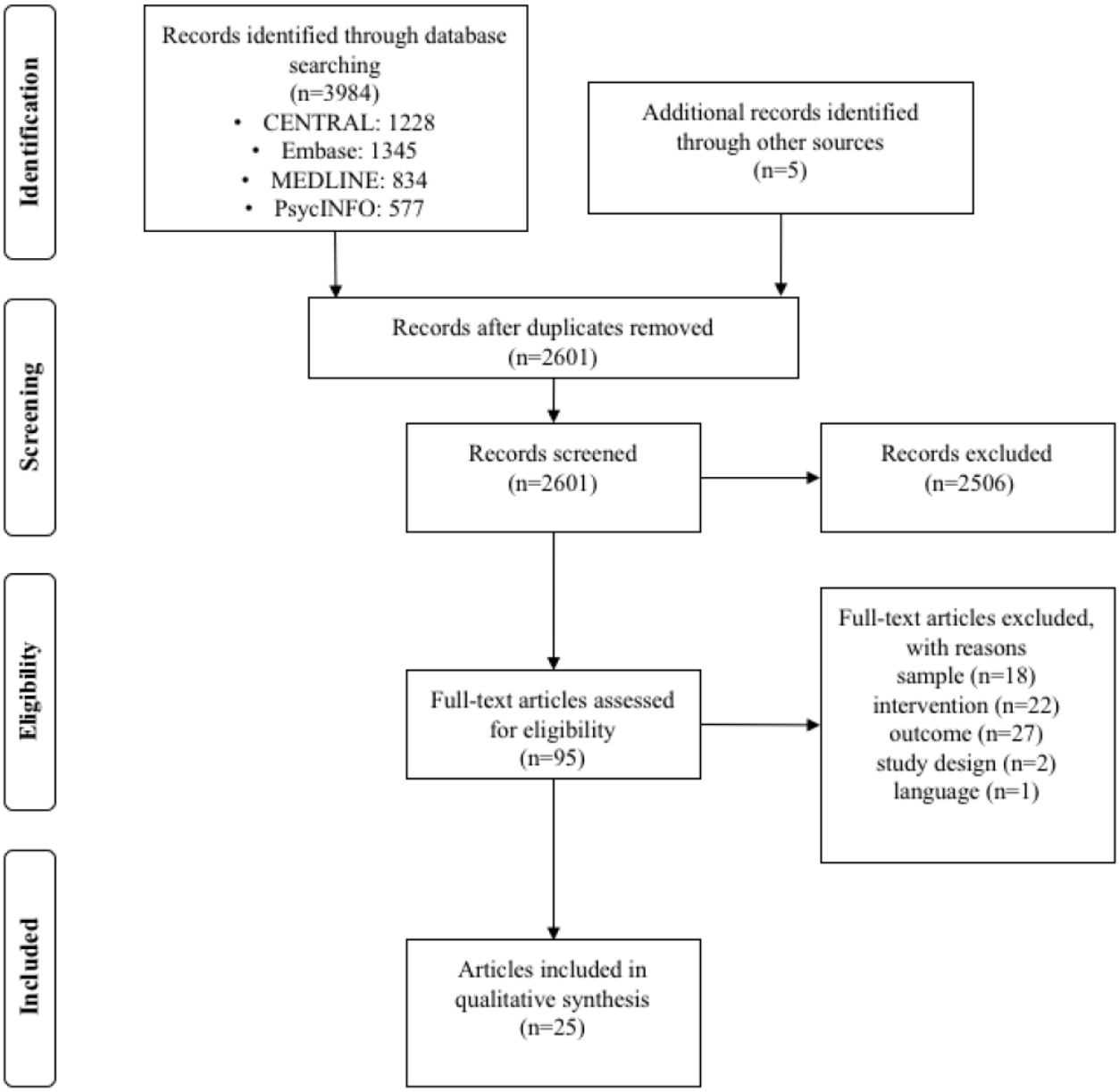
Studies were eligible for inclusion if they fulfilled the following criteria: (1) participants were children (0-13 years) or adolescents (14-21 years) diagnosed with a mental disorder or exhibited clinically relevant symptoms of a mental disorder; (2) in the case of mixed samples with adolescents and young adults, the mean age of the sample was not above 21 years; (3) the interventions were designed for children or adolescents, or for parents of children or adolescents fulfilling the first criterion; (4) the diagnosis of mental disorders was based on the diagnostic and statistical manual of mental disorders, or the International Classification of Diseases criteria and was assessed with a validated and standardized clinical interview, or a standardized self-report instrument, or standardized ratings by significant others (eg, parents, teachers, clinicians); (5) samples of mixed or comorbid mental disorders were included; (6) studies with different recruitment strategies were eligible; (7) interventions with different theoretical orientations were eligible; (8) the intervention was predominantly delivered through the internet (eg, via web browsers or mobile or smartphone apps); (9) interventions with different degrees of human guidance and completely self-guided interventions were eligible for inclusion; (10) different active and passive control groups (CGs) were included; (11) changes in the symptoms and (12) mediators were reported; (13) at least one mediation analysis was performed; (14) studies were RCTs or secondary analyses of RCTs published in a peer reviewed journal in English language.

Systematic Literature Search and Study Selection

The search strategy was 3-fold. First, systematic literature searches were conducted in the Cochrane Central Register of Controlled Trials, Embase, MEDLINE (ie, Ovid MEDLINE, Ovid MEDLINE Epub Ahead of Print, Ovid MEDLINE In-Process and Other Non-Indexed Citations, Ovid MEDLINE Daily Update), and PsycINFO databases from database inception until May 30, 2020. The search strings were built on established prior search strings [24,40] further adapted to specifically meet the research questions at hand and modified for each database in Ovid (for details on all search strings, see Multimedia Appendix 1). Second, the reference lists of all eligible studies and other relevant reviews were manually searched to identify further studies that met our inclusion criteria (ie, backward searches). Third, a citation-search (ie, forward search) was conducted in the Web of Science database.

Study selection was conducted with the support of a software tool for systematic reviews [41]. Duplicates were detected automatically by the software or were manually removed. First, 1 reviewer (SE) screened titles and abstracts of all the remaining studies and discarded irrelevant articles. Second, the full texts of all potentially relevant articles were screened in terms of the aforementioned eligibility criteria independently by 2 reviewers (SE and HN). Disagreements were resolved by consultation with a third reviewer (MD). The full process of the systematic literature search and study selection is displayed in the PRISMA flowchart of Figure 1.

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



Data Extraction

Two independent reviewers (SE and HN) extracted the following data: study information items (name of first author, country, and year), sample information (sample size and age), intervention characteristics (theoretical orientation, number of modules, and duration of intervention), information about control conditions, and information about outcomes (mediator type, instrument, and statistical analysis of mediation). Authors were contacted via email in case of missing information essential for decisions on study selection and the application of the systematic review.

Categorization of Studies

The included studies were divided into 2 categories: studies evaluating interventions with a psychotherapeutic focus (ie, DHI_{PSY}) and studies evaluating interventions with a focus on health promotion (comprising interventions targeting health behavior, lifestyle, and behavior change interventions for the purpose of primary and secondary prevention; ie, DHI_{HP}). Of note, interventions with rehabilitation or tertiary prevention focus were not included.

Quality Assessment

Risk-of-Bias Assessment

The methodological quality of the included studies was assessed independently by 2 reviewers (SE and HN) using the Cochrane risk-of-bias (RoB) tool for randomized trials (version 2, RoB2; [42]) on 5 domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcomes, and (5) bias in selection of the reported result. The included studies were rated as having *low*, *unclear*, or *high RoB* in each domain [42]. Interrater reliability was calculated using the Cohen κ in RStudio (version 1.2.1335; [43]).

Quality Criteria for Process Research and Approximating Causality

The included studies were rated by 2 independent reviewers (SE and HN). The rating system was based on Kazdin's [25] initial criteria to approach causality, modified by Domhardt et al [24] and Lemmens et al [44]. To meet the respective criteria, studies had to: (1) use an appropriate RCT design, (2) include a CG, (3) report a theoretical foundation for mediators, (4) have a minimum sample size of 40 participants per group, (5) examine multiple mediators within 1 study, (6) assess temporality (3 or more assessments of the mediator variables and outcomes), (7) experimentally manipulate the mediator, and (8) reveal a strong statistical association among intervention, mediator, and outcome (operationalized as statistical significance of $P<.05$, as suggested by Moreno-Peral et al [45]). All criteria were rated as fulfilled or not fulfilled. In accordance with prior research [24,44], criteria specificity, consistency, and gradient were not assessed, as they are not meaningfully applicable in single studies (consistency) or are too exclusive for some therapeutic processes (gradient and specificity).

Results

Study Characteristics

Altogether, data from 25 (k=25) publications were analyzed [46-70]. Specifically, 52% (13/25) of the studies [46,47,49,54,56,58-60,64-67,69] evaluated DHI_{PSY} and 48% (12/25) of the studies evaluated DHI_{HP} [48,50-53,55,57,61-63,68,70]. In the studies, a total of 4884 participants were randomized. Studies on DHI_{PSY} accounted for 43.2% (2110/4884) of participants, and studies on DHI_{HP} accounted for 56.8% (2774/4884) of participants. The overall sample sizes varied from 51 [59] to 818 [63]. The median publication year was 2014 (2002 [48] to 2020 [49]). Most studies were conducted in the United States (17/25, 68%), and most of the study participants were female (2985/4884, 61.18%). In 1 study, no information was provided regarding the distribution of sex [54]. The mean age of participants was 18.49 (SD 2.01) years. The lowest mean age was 5.4 (SD 2.2) years [67], and the highest was 21.02 (SD 2.16) years [70]. Participants were younger in studies evaluating DHI_{PSY} compared with participants in studies evaluating DHI_{HP} (17.11 vs 19.67 years). The exact information on the average age of the final sample was not provided in 4 studies [54,56,63,68]. Most interventions took place without parental involvement and were directed at the youth themselves (20/25, 80%). In 20% (5/25) of the studies, interventions for children [54,67] as well as interventions for both children and adolescents [46,65] and for adolescents only [64] were evaluated with the involvement of parents (including intervention components for parents or parent training). Demographic information about the parents who participated was provided in 16% (4/25) of the studies [46,54,65,67]. Study participants were recruited predominantly from educational or health-related settings (15/25, 60%). In 32% (8/25) of the studies, participants were made aware of the study through both web-based advertising (including social media and websites) and conventional advertising (including letters and flyers). To identify potential participants, one study (1/25, 4%) used data from a mass web-based survey [59]. Information on recruitment strategy could not be identified in 4% (1/25) of the studies [68]. Across studies, the average study dropout rate accounted for 18% (range 0% [62] to 39.3% [54]). In 4% (1/25) of the studies, the dropout rate was not reported [46].

The interventions were directed toward a broad range of mental health problems, including risky drinking behavior (including risky or heavy drinking and binge drinking; 11/25, 44%), depressive disorders (5/25, 20%), anxiety disorders (including separation anxiety, generalized anxiety disorder, social phobia, and specific phobia; 3/25, 12%), behavioral problems (2/25, 8%), and insomnia (1/25, 4%). Furthermore, 12% (3/25) of the studies evaluated interventions that addressed multiple disorders simultaneously (ie, transdiagnostic interventions). De Bruin et al [47] addressed transdiagnostic *psychopathological abnormalities* (including affective, anxiety, and somatic problems; problems concerning attention-deficit/hyperactivity disorder; oppositional defiant behavior; and conduct problems). Levin et al [60] addressed *psychological problems* such as depressive disorders, generalized anxiety disorder, social phobia, alcohol consumption, academic worries, worries concerning eating, hostility, and negative stress. Another intervention addressed depression, anxiety, and stress [56]. Of note, there were 2 interventions evaluated and applied in more than 1 study (the *e-CHUG* interventions in 4 studies; and the *BRAVE* interventions in 2 studies), but the investigated mediators differed in all studies; thus, these publications were regarded as distinct entities and single studies (see Multimedia Appendix 2 [46-70] for details).

Internet-based interventions were evaluated in 92% (12/13) of the studies on DHI_{PSY}. Most interventions were based on the components of cognitive behavioral therapy [46,47,49,59,66]. In addition, relaxation strategies, such as progressive muscle relaxation and autogenic training [49], and elements of social learning theory [54] were deployed. Further interventions were based on the acceptance and commitment therapy [60], components of systemic family therapy, problem solving and communication training, cognitive restructuring, and alternative family roles [64]. One intervention [56] was based on the *temporal model of control*. In 23% (3/13) of the studies, precise information on the theoretical orientation and background of the intervention was not available [58,67,69]. Moreover, 7% (1/13) of the studies evaluated a mobile-based intervention based on a self-monitoring program [58].

All DHI_{HP} studies were internet-based and included a feedback component (12/12, 100%). The so-called *e-CHUG* tool was evaluated in 33% (4/12) of the studies [51,57,62,68]. Feedback was provided via email [48,53] or in person using a motivational interviewing approach [51,52]. In addition to the feedback components, cognitive components such as expressive writing [70] and retrieval of remembered information from feedback were evaluated [57]. A detailed overview of the study characteristics is provided in Multimedia Appendix 2.

Mediators

An overview of the classification of mediators and their empirical support is provided in Table 1. A total of 39 potential

mediators were investigated in the included RCTs. Among these, more than half of the mediators were evaluated as significant (21/39, 54%). With 48% (19/39) intervening variables, the largest group of all examined intervening variables was of a cognitive nature. A total of 13 cognitive mediators were evaluated as significant in the primary studies (13/19, 68%). We further divided the group of cognitive mediators into the *assessment* (examined: 8/39, 20%; significant: 5/8), *motivation* (examined: 1/39, 2%; significant: 0/1, 0%) and *cognitive processes* subcategories (examined: 10/39, 25%; significant: 8/10, 80%). Further evenly investigated mediator categories were emotional/affective (examined: 4/39, 10%; significant: 4/4, 100%), interpersonal (examined: 4/39, 10%; significant: 1/4, 25%), and parenting behavior mediators (examined: 4/39, 10%; significant: 0/4, 0%). The second largest group of mediators was not clearly classifiable into one of the aforementioned categories and was thus subsumed into a separate *other* mediator category (examined: 8/39, 20%; significant: 3/8, 38%). Of note, a Wilcoxon rank-sum test revealed no difference in the sample sizes between studies that found at least one significant mediator and studies that found no significant mediator (W {19,6}=59; *P*=.93).

Table 1. Classification of mediators.

Mediators	Studies, n (%) (n=25)	Age (years), range[a]	Guidance	Disorder	Effect size[b]	Significance	Criteria met ≥5
Emotional and affective mediators							
Emotional self-perception	1 (4)	14-22	Unguided	Depression	Partially standardized ES[c]=−1.049 (95% CI −1.35 to −0.755)	Yes	(+)[d]
Fear	1 (4)	16-24	Internet-based psychotherapy	Depression	Between groups ES=0.49 (95% CI 0.24 to 0.75)	Yes	(+)
Hopelessness	1 (4)	>18	Unguided	Depression	Cohen *d*=1.13 (between groups follow-up)	Yes	(+)
Thought-related distress	1 (4)	Undergraduate students	Guided self-help	Generalized anxiety disorder	Cohen *d*=0.526 (between groups follow-up)	Yes	(+)
Interpersonal mediators							
Parent-Youth conflict	1 (4)	12-19	Internet-based psychotherapy	Depression	—[e]	Yes	(+)
Family conflicts related to diabetes management	1 (4)	12-19	Internet-based psychotherapy	Depression	—	No	(+)
Failed help or negative social support	1 (4)	12-19	Internet-based psychotherapy	Depression	—	No	(+)
Social skills	1 (4)	8-17	Guided self-help	Social phobia	—	No	(+)
Parenting behavior mediators							
Appropriate education	1 (4)	10-13	Guided self-help	Behavioral problems	—	No	(+)
Skill in setting clear boundaries	1 (4)	10-13	Guided self-help	Behavioral problems	—	No	(+)
Severity and inconsistent education	1 (4)	10-13	Guided self-help	Behavioral problems	—	No	(+)
Change in parenting behavior	1 (4)	3-9	Guided self-help	Behavioral problems	—	No	(−)[f]
Cognitive mediators							
Assessment							
Assessment of discrepancy[g]	2[g] (8)	College students; College students	Guided self-help; guided self-help	Risky drinking behavior	—; —	No	(+)
Assessment of peer drinking behavior[g]	2[g] (8)	18-24; 18-24	Guided self-help; guided self-help	Risky drinking behavior	—; —	Yes	Different result

Mediators	Studies, n (%) (n=25)	Age (years), range[a]	Guidance	Disorder	Effect size[b]	Significance	Criteria met ≥5
Perceived norm[g]	4[g] (16)	18-25; Students; First semester students; 18-26	Guided self-help; unguided; guided self-help; unguided	Risky drinking behavior	—; Cohen *d*=−0.2 (ES Biannual); —; —	Different results[h]	(+)
Motivation							
Motivation to change drinking behavior	1 (4)	College students	Unguided	Risky drinking behavior	—	No	(+)
Cognitive processes							
Alcohol-related expectations	—	Students	Guided self-help	Risky drinking behavior	—	Yes	(+)
Remembered information	1 (4)	Students	Guided self-help	Risky drinking behavior	—	Yes	(+)
Mastering	1 (4)	16-25	Internet-based psychotherapy	Depression	Between groups ES=0.94 (95% CI 0.64 to 1.23)	Yes	(+)
Willingness to cope	—	Students	Unguided	Depression	—	No	(−)
Cognitive arousal before falling asleep	1 (4)	Students	Unguided	Insomnia	—	No	(+)
Sleep-related cognition	—	Students	Unguided	Insomnia	—	Yes	(+)
Postevent processing	1 (4)	8-17	Guided self-help	Social phobia	—	Yes	(+)
Mindful acceptance	1 (4)	Students	Guided self-help	Transdiagnostic[i]	Proportion mediated ES=range 16.05% to 28.57%	Yes	(−)
Obstruction of appreciation of life	—	Students	Guided self-help	Transdiagnostic[i]	Proportion mediated ES=range 29.18% to 57.94%	Yes	(−)
Perceived control	1 (4)	18-21	Guided self-help	Transdiagnostic[i]	Cohen *d*=0.07; Cohen *d*=0.59; Cohen *d*=0.66 (between groups follow-up)	Yes	(+)
Other							
Alcohol consumption as coping behavior	1 (4)	18-21	Unguided	Risky drinking behavior	—	No	(+)
Therapy adherence	1 (4)	7-18	Guided self-help	Anxiety disorder	—	No	(−)
Eating disorder	1 (4)	18-25	Unguided	Depression	—	Yes	(+)

Mediators	Studies, n (%) (n=25)	Age (years), range[a]	Guidance	Disorder	Effect size[b]	Significance	Criteria met ≥5
Overall sleep quality	1 (4)	Students	Unguided	Insomnia	—	No	(+)
Chronotypical	1 (4)	Students	Unguided	Insomnia	—	No	(+)
Physical arousal before falling asleep	1 (4)	Students	Unguided	Insomnia	—	Yes	(+)
Trauma-related sleeping disorder	1 (4)	Students	Unguided	Insomnia	—	No	(+)
Insomnia	1 (4)	12-19	Guided self-help	Transdiagnostic[i]	—	Yes	(+)

[a]If age range was not reported, participant group labels were used.

[b]Effect size measures differed across studies.

[c]ES: effect size

[d]Met 5 or more criteria.

[e]Not available.

[f]Met fewer than 5 criteria.

[g]If mediator was assessed in more than 1 study, data and results were separated with ";".

[h]The only mediator nonsex-specific perceived norm was not significant.

[i]Transdiagnostic intervention targets more than one disorder.

A broad range of different approaches to mediation analyses was deployed, with some studies relying on several strategies at once. Bootstrapping (eg, Preacher and Hayes [71]) was used in almost half of the studies (11/25, 44%). In 32% (8/25) of the studies, the mediation analysis was based on the approach of Baron and Kenny [30], and in addition, 8% (2/25) performed a Sobel test. Furthermore, 32% (8/25) of the mediation analyses were performed using structural equation modeling (7/25, 28% using the MPlus software). Moreover, 4% (1/25) of the studies performed multiple regression analysis according to the Judd and colleagues paradigm [72], and another study did not provide any information on the statistical analysis.

Quality Assessment of Included Studies

Risk-of-Bias Assessment

As illustrated in the *RoB2 Graph* (Figure 2), approximately three-quarters (19/25, 76%) of the included studies were rated as having a *high* RoB. In more than half of the studies (14/25, 56%), the *bias due to deviation from the intended intervention* was rated as *high*. The RoB on this domain was more often evaluated to be *high* in studies on DHI_{HP} (10/25, 40%) when compared with studies on DHI_{PSY} (4/25, 16%). Both the randomization procedure and the process of reporting results were predominantly rated with *some concern* across studies (19/25, 76% and 22/25, 88%). Interrater reliability varied across domains from Cohen κ=0.76 to Cohen κ=0.93. According to Landis and Koch [73], these agreements can be rated as *substantial* to *almost perfect*. A summary of the RoB2 assessments is shown in Figure 3.

Figure 2. Risk-of-bias graph.

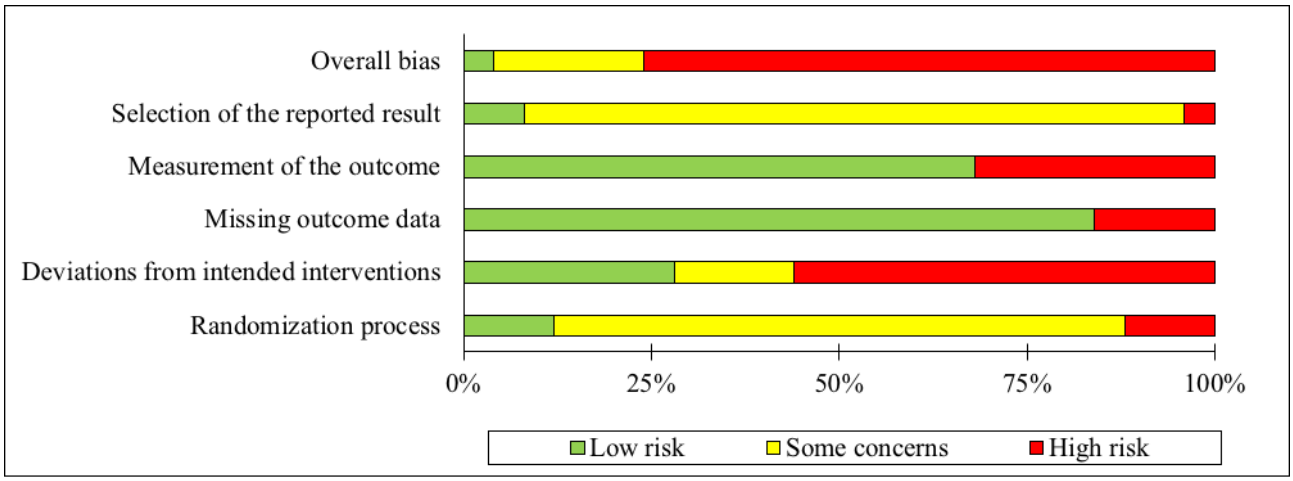
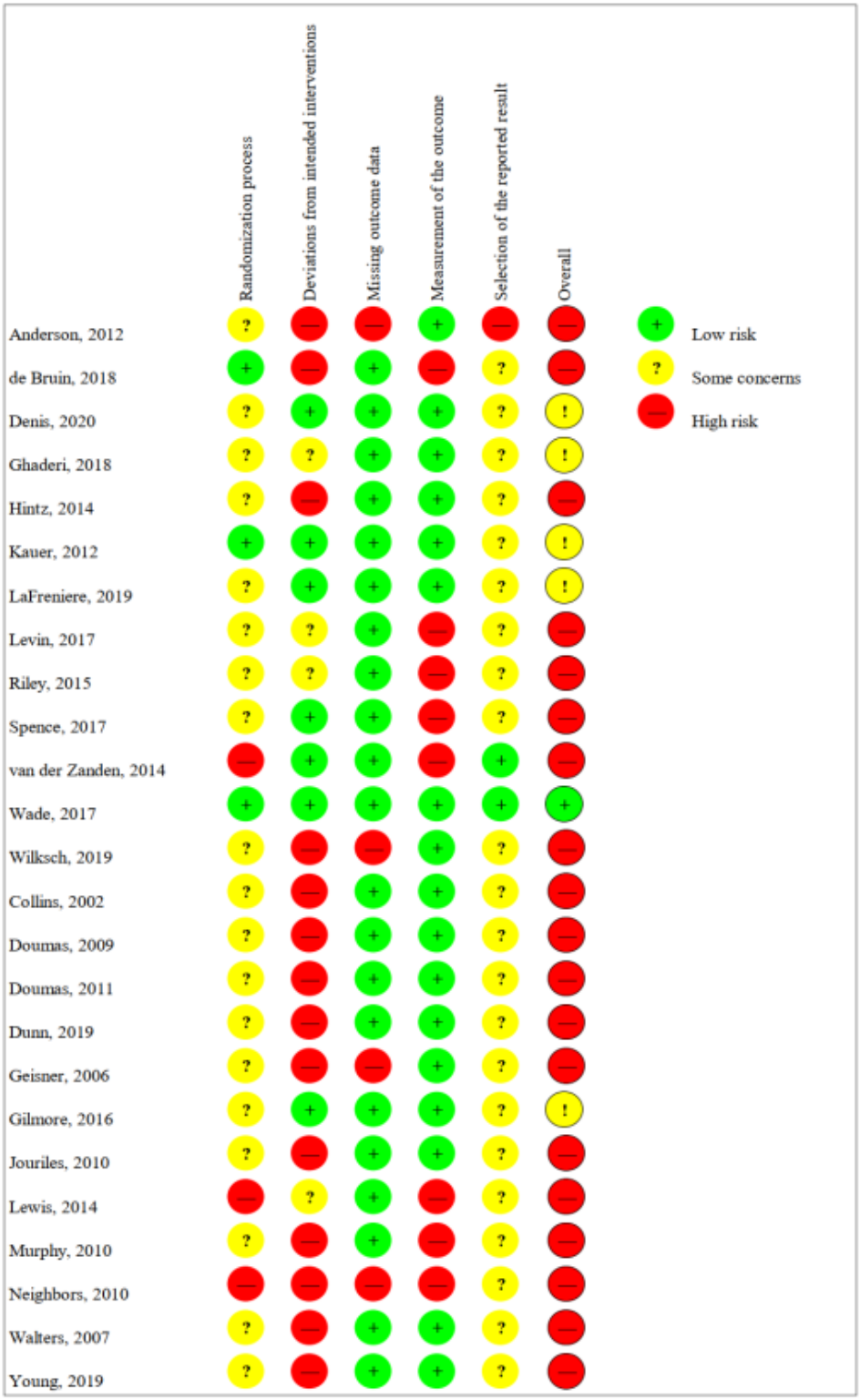


Figure 3. Risk-of-bias summary.



Evaluation of the Quality Criteria for Process Research and Approximating Causality

The evaluation of the included studies with regard to the methodological quality of process research revealed that in both the DHI_{PSY} and the DHI_{HP} groups, most studies (DHI_{PSY}: 9/13, 69% vs DHI_{HP}: 11/12, 92%) fulfilled 5 or more out of the 8 criteria. Owing to the eligibility criteria used in this review, almost all studies met the requirement of an RCT design (24/25, 96%) and a CG (24/25, 96%). In the publication by Anderson et al [46], the results from 8% (2/25) of the studies were jointly reported, with only the second study evaluating a DHI_{PSY} without a direct comparison with the CG, although this criterion was fulfilled in the first study. Even if mediators were collected at more than 2 measurement time points in more than half of the studies (15/25, 60%; including follow-up), an evaluation of the chronology of changes in the mediator variable or variables and outcomes was conducted in only 4% (1/25; de Bruin et al

[47]) of the studies. Furthermore, 8% (2/25) of the studies, Hintz et al [56] and Jouriles et al [57] implemented direct experimental manipulation of mediators. A detailed overview of the evaluation of the methodological quality criteria for process research and the approximation of causality are outlined in Tables 2 and 3, as well as in Multimedia Appendix 3.

Table 2. Quality criteria for process research and approximation of causality (n=25).

Studies		Randomized controlled trial	CG[a]	Theoretical foundation	n≥40 (Each CG and EG[b])	Various mediators	Time sequence	Manipulation	*P* value (<.05)[c]	Σ (Yes)
DHI_{PSY}[d]										
	Anderson, 2012 [46][e]	No[f]	No	Yes[g]	No	No	No	No	Yes	2
	de Bruin et al [47]	Yes	Yes	Yes	No	No	Yes	No	Yes	5
	Denis et al [49]	Yes	Yes	No	Yes	Yes	No[h]	No	Yes	5
	Ghaderi et al [54]	Yes	Yes	Yes[i]	Yes	Yes	No[h]	No	Yes	6
	Hintz, 2014 [56]	Yes	Yes	Yes	Yes	No	No[h]	Yes	Yes	6
	Kauer et al [58]	Yes	Yes	Yes	Yes	No	No[h]	No	Yes	5
	LaFreniere, and Newman [59]	Yes	Yes	Yes	No	No	No[h]	No	Yes	4
	Levin et al [60]	Yes	Yes	No	No	Yes	No	No	Yes	4
	Riley et al [64]	Yes[j]	Yes[j]	Yes	Yes	Yes	No[h]	No	Yes	6
	Spence et al [65]	Yes	Yes	Yes	No	Yes	No[h,k]	No	Yes	5
	Zanden et al, 2014 [66]	Yes	Yes	Yes	Yes	Yes	No[h]	No	Yes	6
	Wade et al [67]	Yes	Yes	Yes	No[i]	No	No[h]	No	Yes	4
	Wilksch, O'Shea, and Wade [69]	Yes	Yes	Yes	Yes	No	No[h]	No	Yes	5
DHI_{HP}[l]										
	Collins, Carey, and Sliwinski [48]	Yes	Yes	Yes	Yes	No	No[h]	No	Yes	5
	Doumas, McKinley, and Book [50]	Yes	Yes	Yes	No	No	No	No	Yes	4
	Doumas et al [51]	Yes	Yes	Yes	Yes	No	No	No	Yes	5
	Dunn, 2019 [52]	Yes	Yes	Yes	Yes	No	No	No	Yes	5
	Geisner, Neighbors, and Larimer [53]	Yes	Yes	Yes	Yes	Yes	No	No	Yes	6

Studies		Randomized controlled trial	CG[a]	Theoretical foundation	n≥40 (Each CG and EG[b])	Various mediators	Time sequence	Manipulation	*P* value (<.05)[c]	Σ (Yes)
	Gilmore and Bountress [55]	Yes	Yes	Yes	Yes	No	No	No	Yes	5
	Jouriles et al [57]	Yes	Yes	Yes	No	No	No	Yes	Yes	5
	Lewis et al [61]	Yes	Yes	Yes	Yes	No	No[h]	No	Yes	5
	Murphy, 2010 [62][h]	Yes	Yes	Yes	Yes	Yes	No[h]	No	Yes	6
	Neighbors et al [63]	Yes	Yes	Yes	Yes	Yes	No[h]	No	Yes	6
	Walters, Vader, and Harris [68]	Yes	Yes	Yes	Yes[m]	No	No[h]	No	Yes	5
	Young and Neighbors [70]	Yes	Yes	Yes	Yes	No	No	No	Yes	5

[a]CG: control group.

[b]EG: experimental group.

[c]Overall significance level *P*<.05; only data from study 2 taken into account.

[d]DHI_{PSY}: digital health interventions with a psychotherapeutic focus.

[e]Only data from study 2 are taken into account.

[f]*No* indicates criteria not met.

[g]*Yes* indicates criteria met.

[h]More than 2 measurements (including follow-up) reported, but no evaluation of time sequence.

[i]Subscales have no theoretical foundations.

[j]Initial study had a randomized controlled trial design; in secondary analysis, both groups were taken together.

[k]Due to missing follow-up data in the waitlist condition, mediation analysis was conducted only with data from baseline and after 12 weeks (at least 3 sessions were completed).

[l]DHI_{HP}: digital health interventions with a focus on health promotion.

[m]Information was given after contacting authors.

Table 3. Number of studies meeting each single quality criterion for process research (n=25).

Criterion	DHI_{PSY}[a], n (%)	DHI_{HP}[b], n (%)	Overall, n (%)
Randomized controlled trial	12 (48)	12 (48)	24 (96)
CG[c]	12 (48)	12 (48)	24 (96)
Theoretical foundation	11 (44)	12 (48)	23 (92)
n≥40 (for CG and EG[d] each)	7 (28)	10 (40)	17 (68)
Evaluation of various mediators	6 (24)	3 (12)	9 (36)
Time sequence or temporality	1 (4)	0 (0)	1 (4)
Manipulation of mediators	2 (8)	0 (0)	2 (8)
P<.05	13 (52)	12 (48)	25 (100)

[a]DHI_{PSY}: digital health interventions with a psychotherapeutic focus.

[b]DHI_{HP}: digital health interventions with a focus on health promotion.

[c]CG: control group.

[d]EG: experimental group.

Discussion

Principal Findings

This systematic review, to the knowledge of the authors the first of its kind, comprehensively evaluated research on mediators and mechanisms of change in DHIs for common mental disorders in youth. Altogether, 25 studies were included in the review, that have examined 39 distinct mediators. Cognitive variables were found to be the most often investigated mediators, followed by a broad range of other mediators. Even though our eligibility criteria were not limited to a specific mental health condition in youth, only a rather low number of studies were identified by our systematic literature searches, a finding that corresponds to the limited evidence base of research on mechanisms of change in psychotherapy in general [37,74], and for children and adolescents in specific [24,35,37], necessitating further high-quality research efforts to improve interventions and mental health care practices for this younger age group.

Remarkably, our findings indicate that the mediator category with the highest percentage of significant intervening variables was the affective or emotional mediator group (4/4, 100%). The proportions of significance in other mediator categories were by far less high and included combined cognitive (13/19, 68%), other (3/8, 37%), interpersonal (1/4, 25%), and parenting behavior–related mediators (0/4, 0%). The consistent nonsignificant findings on parenting behavior–related intervening variables are astonishing, as parenting behavior is otherwise thought to be of paramount importance in the treatment of behavioral problems in youth, both in conventional interventions delivered face-to-face [75] and in digital parent training alike [76,77]. Therefore, the findings of our study are in contrast to those of a systematic review [75], which revealed that in 45% (39/86) of the included studies on face-to-face parent training programs, parenting behavior served as a mediator for the association between the intervention and symptom change in children. Furthermore, the importance of emotion regulation might be underestimated in (digital) psychotherapeutic interventions for children and adolescents, considering the consistent pattern of significant findings across studies in our review, as well as by allowing for comprehensive earlier research highlighting the overall importance of emotion regulation competencies for mental health in childhood and adolescence [78]. However, given the rather small number of mediation studies per category in our review, these findings need to be interpreted with caution and must be considered as preliminary. This is also owing to the rather heterogeneous evidence base, where included studies varied broadly in terms of the theoretical foundations of the intervention as well as the simultaneous consideration of various mental disorders, which may further restrict the comparability between studies. Nevertheless, a Wilcoxon rank-sum test revealed no differences in the sample sizes between studies that found at least one significant mediator and studies that found no significant mediator, suggesting that there is no effect of sample size on the evaluation of intervening variables, strengthening the robustness and validity of the findings on the mediators in this regard.

Importantly, participants were most often adolescents, with only 16% (4/25) of eligible studies evaluating interventions for children, suggesting an additional research gap for this younger age group. Here, all interventions for children were conducted with the involvement of parents, suggesting a crucial role of the accompanying human support in DHIs for younger children, which must be corroborated and further specified by forthcoming research [12,79]. In studies evaluating DHI_{HP}, only cognitive mediators (perceived norm and alcohol-related expectation) were investigated. This finding is consistent with prior research, where the cognitive mediator, perceived norm, was established as one of the most evaluated mediators in interventions for problematic drinking behaviors in adolescents and young adults delivered on site [80]. However, this result rather highlights the researcher's presumptions than the evidence base of the superior relevance of cognitive mediators over other, not yet well examined, affective and behavioral mediators.

The results of the RoB assessment indicated a rather limited overall study quality, with 76% (19/25) of the included studies rated with a *high RoB*, a finding that aligns with the review by Christ et al [81], in which 92% (22/24) included studies on DHIs for adolescents and young adults were assessed with a *high RoB*. These findings on RoB2 might be largely due to the specifics of psychotherapy research [82], where masking of therapists or personnel and participants is difficult to achieve, as well as the more conservative novel RoB2 algorithm [83]. Of note, in our review, the mean study dropout rate was 18%. Therefore, only a fraction of the included studies fulfilled the RoB2 criterion that 95% of randomized participants' data should be available for data analysis. Although the current mean dropout rate in this review is rather small compared with other high dropout rates found in DHIs, the well-known issue of limited engagement still warrants future research to further investigate approaches to remedy (study and intervention) attrition [84,85]. In this particular field, primary studies in the review at hand may offer guidance and direction to more effective engagement in youth (Multimedia Appendix 2), such as interactive intervention components and age-appropriate content presentation in the form of puzzles, videos, or cartoons [12].

The results of the methodological quality assessment for process research revealed that most studies adhered to certain quality criteria satisfactorily (ie, using an RCT design and CG, evaluation of a strong statistical association, describing a theoretical foundation for mediators and sample size per trial group). However, in contrast, most primary studies did not experimentally manipulate the potential mediators, did not assess several mediators simultaneously, and did not evaluate the time sequence or temporal ordering of changes in mediators and outcomes. However, this latter criterion is considered to be of utmost importance for causal inferences and is sometimes referred to as the *fifth step* of mediation analysis [86]. Even though mediators and outcomes were assessed more than twice in 60% (15/25) of the included studies, only 4% (1/25) of the studies [47] actually conducted a statistical evaluation of the time sequence of cause and effect. Taken together, the findings of our review seem to be in line with those of other research [24,44,45,83], pointing to important methodological

shortcomings of mediation studies, which should be amended by future research. At this point, 1 high-quality RCT [58] sets an example, fulfilling 6 out of 8 quality criteria.

Strengths and Limitations

This systematic review offers several strengths, including a comprehensive literature search, as well as the broad consideration of common mental disorders and theoretical orientations, resulting in an extensive overview of the research on mediators and mechanisms of change for DHIs in children and adolescents conducted so far. Furthermore, with the differential consideration of the 2 categories of DHI_{PSY} and DHI_{HP}, we intended to provide specific evidence-based information that might be valuable for digital psychotherapeutic and health promotion interventions alike. However, some limitations must be considered when interpreting the findings of this review. First, the number of eligible studies is rather small; the generalizability of the findings might be reduced as most studies were conducted in Western countries; these studies evaluated internet-based interventions; participants were predominantly female; and these studies relied on older participants. Second, only English language papers were included, and we did not incorporate gray literature in our study as recent findings indicated a negligible relevance of gray literature in systematic reviews [87]; further, publication bias cannot be ruled out. Third, although we only included RCT studies in our review (which might have led to an overestimation of the study quality in general, omitting other design studies with potentially lower methodological rigor), the RoB assessment indicated substantial shortcomings in the included studies. However, this finding should be weighed against the typical constraints of research on psychological interventions.

Future Directions

The following recommendations might be helpful for future mediation studies, aiming to advance our understanding of the working mechanisms of DHIs for youth. First, it is essential to avoid the methodological shortcomings outlined in the assessment of quality criteria for process research, especially with regard to the temporality of changes and experimental manipulation of mediators, both of which are key for the justification of causal inferences. Furthermore, future studies should resort to more sophisticated and current methods of mediation analysis (eg, Grimm et al [88] and Hofmann et al [89]), ideally capable of clarifying temporal precedence and illustrating the actual pattern of change [24]. Second, the therapeutic or working alliance (ie, the professional relationship between therapists and patients; eg, Grawe et al [90]) was not evaluated as a mediator in any of the included studies. However, the recent development of the working alliance inventory for digital interventions [91] may contribute to future process research, informing the evidence base of the effects of a *digital* therapeutic alliance in DHIs for youth [37,92]. Furthermore, mediators that have not been investigated so far, such as behavioral and biological variables, should be evaluated in future studies. At this point, the possibilities of novel technological methods in process research seem not to be fully exploited in this research field [37]. Passive sensing methodologies and digital phenotyping approaches with smartphones [93] might generate insights on behavioral and biological mechanisms in real life, in addition to the predominant information on cognitive variables derived from self-reports identified in the current review. Third, although worldwide about 90% of youth live in countries with low or medium income, 90% of RCTs investigating mental disorders in children and adolescents are conducted in high-income countries [1], mirroring the findings of the review at hand. In contrast, the presumed advantages of DHIs may be especially relevant in structurally weak and low-income countries. To overcome this contradiction, replication studies should aim to include and reach populations from low-income countries as well. Finally, our review revealed that mediators focusing on emotional and cognitive processes might be of paramount importance as potential mechanisms of change in DHIs for youth. However, the processes covered by these mediators varied to some extent. Thus, future studies focusing on both emotional and cognitive processes in a systematic way within one framework could be of great importance for the field.

Conclusions

This systematic review is the first to comprehensively investigate the mechanisms of change in different DHIs for youth. The key findings indicate that the largest group of examined mediators are cognitive variables, followed by an array of other mediator variables, including interpersonal, parenting behavior, and affective mediator categories. Of note, emotional and affective mediators consistently reached statistical significance across studies, whereas parenting-related mediators were evaluated constantly as nonsignificant. However, these findings must be considered cautiously, as we detected only a limited number of primary studies, despite including a broad range of mental disorders and interventions. Future studies should aim to lessen this research gap, ideally adhering to the quality criteria for process research and recent methods of mediation analyses, enabling more causally robust findings. These forthcoming studies might contribute to disentangling the therapeutic processes in DHIs, providing evidence-based knowledge to inform intervention development and augmented mental health care practices worldwide.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (grant identification FKZ01KG1802). The funding source, the Federal Ministry of Education and Research, had no role in the study design, collection, analysis, or interpretation of the data, writing of the manuscript, or the decision to submit this paper for publication.

Authors' Contributions

MD, SE, and HB developed the study design. SE and HN conducted the systematic literature searches, extracted the data, and rated the risk of bias and methodological quality criteria of included studies under the supervision of MD. MD wrote the first draft of the manuscript. All authors (MD, SE, HN, AL, AB, and HB) have contributed to the writing of the manuscript and have approved the final manuscript.

Conflicts of Interest

AB has received payment for consulting from Pro-Change Behavior Systems. All other authors (MD, SE, HN, AL, and HB) stated no conflicts of interest.

Multimedia Appendix 1
Search strings for Cochrane Central Register of Controlled Trials, Embase, MEDLINE, and PsycINFO in Ovid.
[DOCX File , 64 KB - jmir_v23i11e29742_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 113 KB - jmir_v23i11e29742_app2.docx]

Multimedia Appendix 3
Studies with number of criteria met.
[DOCX File , 68 KB - jmir_v23i11e29742_app3.docx]

References

1. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-1525. [doi: 10.1016/S0140-6736(11)60827-1] [Medline: 22008427]
2. Erskine HE, Moffitt TE, Copeland WE, Costello EJ, Ferrari AJ, Patton G, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. Psychol Med 2015 May;45(7):1551-1563 [FREE Full text] [doi: 10.1017/S0033291714002888] [Medline: 25534496]
3. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
4. Costello EJ, He J, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with psychiatric disorders: 12-month data from the National Comorbidity Survey-Adolescent. Psychiatr Serv 2014 Mar 01;65(3):359-366 [FREE Full text] [doi: 10.1176/appi.ps.201100518] [Medline: 24233052]
5. Jörg F, Visser E, Ormel J, Reijneveld SA, Hartman CA, Oldehinkel AJ. Mental health care use in adolescents with and without mental disorders. Eur Child Adolesc Psychiatry 2016 May;25(5):501-508 [FREE Full text] [doi: 10.1007/s00787-015-0754-9] [Medline: 26275772]
6. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019 Apr 01;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
7. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
8. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
9. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
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. United Nations Children's Fund (UNICEF). The State of the World's Children 2017 Children in a Digital World. New York, NY: United Nations; 2018.
12. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]

13. Lin J, Ebert DD, Lehr D, Berking M, Baumeister H. [Internet based cognitive behavioral interventions: state of the art and implementation possibilities in rehabilitation]. Rehabilitation (Stuttg) 2013 Jun;52(3):155-163. [doi: 10.1055/s-0033-1343491] [Medline: 23761204]
14. Lin J, Baumeister H. Internet- und mobilebasierte interventionen in der psychotherapie. Public Health Forum 2015 Sep 1;23(3):176-179. [doi: 10.1515/pubhef-2015-0063]
15. 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]
16. Ebert DD, Zarski A, 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 Mar 18;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
17. 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 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
18. 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 Rat Emo Cognitive Behav Ther 2015 Nov 17;34(1):31-50. [doi: 10.1007/s10942-015-0228-5]
19. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: a systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
20. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
21. 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]
22. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: a meta-analytic review of intervention components. Depress Anxiety 2019 Mar;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
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. Domhardt M, Steubl L, Boettcher J, Buntrock C, Karyotaki E, Ebert DD, et al. Mediators and mechanisms of change in internet- and mobile-based interventions for depression: a systematic review. Clin Psychol Rev 2021 Feb;83:101953. [doi: 10.1016/j.cpr.2020.101953] [Medline: 33422841]
25. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. Annu Rev Clin Psychol 2007;3:1-27. [doi: 10.1146/annurev.clinpsy.3.022806.091432] [Medline: 17716046]
26. Kazdin AE, Nock MK. Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. J Child Psychol Psychiatry 2003 Nov;44(8):1116-1129. [doi: 10.1111/1469-7610.00195] [Medline: 14626454]
27. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593-614 [FREE Full text] [doi: 10.1146/annurev.psych.58.110405.085542] [Medline: 16968208]
28. von Soest T, Hagtvet KA. Mediation analysis in a latent growth curve modeling framework. Struct Equ Modeling 2011 Apr 06;18(2):289-314. [doi: 10.1080/10705511.2011.557344]
29. Iacobucci D, Saldanha N, Deng X. A meditation on mediation: evidence that structural equations models perform better than regressions. J Consum Psychol 2007 Apr;17(2):139-153. [doi: 10.1016/s1057-7408(07)70020-7]
30. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
31. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [doi: 10.1001/archpsyc.59.10.877] [Medline: 12365874]
32. Kazdin AE. Psychotherapy for children and adolescents. Annu Rev Psychol 2003;54:253-276. [doi: 10.1146/annurev.psych.54.101601.145105] [Medline: 12185210]
33. Weersing VR, Weisz JR. Mechanisms of action in youth psychotherapy. J Child Psychol Psychiatry 2002 Jan;43(1):3-29. [doi: 10.1111/1469-7610.00002] [Medline: 11848335]
34. Ng MY, DiVasto KA, Cootner S, Gonzalez N, Weisz JR. What do 30 years of randomized trials tell us about how psychotherapy improves youth depression? A systematic review of candidate mediators. Clin Psychol Sci Pract 2020 Aug 18:- (forthcoming). [doi: 10.1111/cpsp.12367]
35. Schmidt SJ, Schimmelmann BG. Mechanisms of change in psychotherapy for children and adolescents: current state, clinical implications, and methodological and conceptual recommendations for mediation analysis. Eur Child Adolesc Psychiatry 2015 Mar;24(3):249-253. [doi: 10.1007/s00787-015-0698-0] [Medline: 25711288]

36. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
37. Domhardt M, Cuijpers P, Ebert DD, Baumeister H. More light? Opportunities and pitfalls in digitalized psychotherapy process research. Front Psychol 2021 Mar 19;12:544129 [FREE Full text] [doi: 10.3389/fpsyg.2021.544129] [Medline: 33815184]
38. 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]
39. Mediators and mechanisms of change in digital health interventions for mental disorders in children and adolescents. Center for Open Science. URL: https://osf.io/vds4e/ [accessed 2021-11-09]
40. Thabrew H, Stasiak K, Hetrick S, Wong S, Huss J, Merry S. E-Health interventions for anxiety and depression in children and adolescents with long-term physical conditions. Cochrane Database Syst Rev 2018 Aug 15;8(8):CD012489 [FREE Full text] [doi: 10.1002/14651858.CD012489.pub2] [Medline: 30110718]
41. Covidence systematic review management. Brown University. URL: https://libguides.brown.edu/Reviews/Covidence [accessed 2021-10-24]
42. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
43. RStudio: integrated development environment for R. R Studio. URL: https://www.rstudio.com/products/rstudio/#:~:text=RStudio%20is%20an%20integrated%20development,history%2C%20debugging%20and%20workspace%20management [accessed 2021-10-24]
44. Lemmens LH, Müller VN, Arntz A, Huibers MJ. Mechanisms of change in psychotherapy for depression: an empirical update and evaluation of research aimed at identifying psychological mediators. Clin Psychol Rev 2016 Dec;50:95-107. [doi: 10.1016/j.cpr.2016.09.004] [Medline: 27770716]
45. Moreno-Peral P, Bellón J, Huibers MJ, Mestre JM, García-López LJ, Taubner S, et al. Mediators in psychological and psychoeducational interventions for the prevention of depression and anxiety. A systematic review. Clin Psychol Rev 2020 Mar;76:101813. [doi: 10.1016/j.cpr.2020.101813] [Medline: 32045780]
46. Anderson RE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: comparison with clinic delivery and its role in predicting outcome. J Med Internet Res 2012 Jun 28;14(3):e88 [FREE Full text] [doi: 10.2196/jmir.1848] [Medline: 22789657]
47. de Bruin EJ, Bögels SM, Oort FJ, Meijer AM. Improvements of adolescent psychopathology after insomnia treatment: results from a randomized controlled trial over 1 year. J Child Psychol Psychiatry 2018 May;59(5):509-522. [doi: 10.1111/jcpp.12834] [Medline: 29052846]
48. Collins SE, Carey KB, Sliwinski MJ. Mailed personalized normative feedback as a brief intervention for at-risk college drinkers. J Stud Alcohol 2002 Sep;63(5):559-567. [doi: 10.15288/jsa.2002.63.559] [Medline: 12380852]
49. Denis D, Eley TC, Rijsdijk F, Zavos HM, Keers R, Espie CA, et al. Is digital cognitive behavioural therapy for insomnia effective in treating sub-threshold insomnia: a pilot RCT. Sleep Med 2020 Feb;66:174-183. [doi: 10.1016/j.sleep.2019.10.007] [Medline: 31901759]
50. Doumas DM, McKinley LL, Book P. Evaluation of two Web-based alcohol interventions for mandated college students. J Subst Abuse Treat 2009 Jan;36(1):65-74. [doi: 10.1016/j.jsat.2008.05.009] [Medline: 18657941]
51. Doumas DM, Workman C, Smith D, Navarro A. Reducing high-risk drinking in mandated college students: evaluation of two personalized normative feedback interventions. J Subst Abuse Treat 2011 Jun;40(4):376-385. [doi: 10.1016/j.jsat.2010.12.006] [Medline: 21295938]
52. Dunn ME, Fried-Somerstein A, Flori JN, Hall TV, Dvorak RD. Reducing alcohol use in mandated college students: a comparison of a Brief Motivational Intervention (BMI) and the Expectancy Challenge Alcohol Literacy Curriculum (ECALC). Exp Clin Psychopharmacol 2020 Feb;28(1):87-98. [doi: 10.1037/pha0000290] [Medline: 31008641]
53. Geisner IM, Neighbors C, Larimer ME. A randomized clinical trial of a brief, mailed intervention for symptoms of depression. J Consult Clin Psychol 2006 Apr;74(2):393-399. [doi: 10.1037/0022-006X.74.2.393] [Medline: 16649884]
54. Ghaderi A, Kadesjö C, Björnsdotter A, Enebrink P. Randomized effectiveness Trial of the Family Check-Up versus Internet-delivered Parent Training (iComet) for families of children with conduct problems. Sci Rep 2018 Jul 31;8(1):11486 [FREE Full text] [doi: 10.1038/s41598-018-29550-z] [Medline: 30065246]
55. Gilmore AK, Bountress KE. Reducing drinking to cope among heavy episodic drinking college women: secondary outcomes of a web-based combined alcohol use and sexual assault risk reduction intervention. Addict Behav 2016 Oct;61:104-111 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.007] [Medline: 27262965]
56. Hintz S, Frazier PA, Meredith L. Evaluating an online stress management intervention for college students. J Couns Psychol 2015 Apr;62(2):137-147. [doi: 10.1037/cou0000014] [Medline: 24635586]
57. Jouriles EN, Brown AS, Rosenfield D, McDonald R, Croft K, Leahy MM, et al. Improving the effectiveness of computer-delivered personalized drinking feedback interventions for college students. Psychol Addict Behav 2010 Dec;24(4):592-599. [doi: 10.1037/a0020830] [Medline: 21198222]

58. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
59. LaFreniere LS, Newman MG. The impact of uncontrollability beliefs and thought-related distress on ecological momentary interventions for generalized anxiety disorder: a moderated mediation model. J Anxiety Disord 2019 Aug;66:102113 [FREE Full text] [doi: 10.1016/j.janxdis.2019.102113] [Medline: 31362145]
60. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
61. Lewis MA, Patrick ME, Litt DM, Atkins DC, Kim T, Blayney JA, et al. Randomized controlled trial of a web-delivered personalized normative feedback intervention to reduce alcohol-related risky sexual behavior among college students. J Consult Clin Psychol 2014 Jun;82(3):429-440 [FREE Full text] [doi: 10.1037/a0035550] [Medline: 24491076]
62. Murphy JG, Dennhardt AA, Skidmore JR, Martens MP, McDevitt-Murphy ME. Computerized versus motivational interviewing alcohol interventions: impact on discrepancy, motivation, and drinking. Psychol Addict Behav 2010 Dec;24(4):628-639 [FREE Full text] [doi: 10.1037/a0021347] [Medline: 21198224]
63. Neighbors C, Lewis MA, Atkins DC, Jensen MM, Walter T, Fossos N, et al. Efficacy of web-based personalized normative feedback: a two-year randomized controlled trial. J Consult Clin Psychol 2010 Dec;78(6):898-911 [FREE Full text] [doi: 10.1037/a0020766] [Medline: 20873892]
64. Riley AR, Duke DC, Freeman KA, Hood KK, Harris MA. Depressive symptoms in a trial behavioral family systems therapy for diabetes: a post hoc analysis of change. Diabetes Care 2015 Aug;38(8):1435-1440. [doi: 10.2337/dc14-2519] [Medline: 26015558]
65. Spence SH, Donovan CL, March S, Kenardy JA, Hearn CS. Generic versus disorder specific cognitive behavior therapy for social anxiety disorder in youth: a randomized controlled trial using internet delivery. Behav Res Ther 2017 Mar;90:41-57. [doi: 10.1016/j.brat.2016.12.003] [Medline: 27988427]
66. van der Zanden R, Galindo-Garre F, Curie K, Kramer J, Cuijpers P. Online cognitive-based intervention for depression: exploring possible circularity in mechanisms of change. Psychol Med 2014 Apr;44(6):1159-1170. [doi: 10.1017/S003329171300175X] [Medline: 23866103]
67. Wade SL, Cassedy AE, Shultz EL, Zang H, Zhang N, Kirkwood MW, et al. Randomized clinical trial of online parent training for behavior problems after early brain injury. J Am Acad Child Adolesc Psychiatry 2017 Nov;56(11):930-9.e2. [doi: 10.1016/j.jaac.2017.09.413] [Medline: 29096775]
68. Walters ST, Vader AM, Harris TR. A controlled trial of web-based feedback for heavy drinking college students. Prev Sci 2007 Mar;8(1):83-88. [doi: 10.1007/s11121-006-0059-9] [Medline: 17136461]
69. Wilksch SM, O'Shea A, Wade TD. Depressive symptoms, alcohol and other drug use, and suicide risk: prevention and treatment effects from a two-country online eating disorder risk reduction trial. Int J Eat Disord 2019 Feb;52(2):132-141. [doi: 10.1002/eat.23005] [Medline: 30582194]
70. Young CM, Neighbors C. Incorporating writing into a personalized normative feedback intervention to reduce problem drinking among college students. Alcohol Clin Exp Res 2019 May;43(5):916-926 [FREE Full text] [doi: 10.1111/acer.13995] [Medline: 30817010]
71. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/brm.40.3.879] [Medline: 18697684]
72. Judd CM, Kenny DA, McClelland GH. Estimating and testing mediation and moderation in within-subject designs. Psychol Methods 2001 Jun;6(2):115-134. [doi: 10.1037/1082-989x.6.2.115] [Medline: 11411437]
73. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
74. Cuijpers P, Reijnders M, Huibers MJ. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019 May 07;15:207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
75. Forehand R, Lafko N, Parent J, Burt KB. Is parenting the mediator of change in behavioral parent training for externalizing problems of youth? Clin Psychol Rev 2014 Dec;34(8):608-619 [FREE Full text] [doi: 10.1016/j.cpr.2014.10.001] [Medline: 25455625]
76. 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]
77. Baumel A, Mathur N, Pawar A, Muench F. Psychosocial interventions for children with externalized behavior problems: an updated meta-analysis of moderator effects. J Child Fam Stud 2021 Jan 07;30(1):65-86. [doi: 10.1007/s10826-020-01863-6]
78. Compas BE, Jaser SS, Bettis AH, Watson KH, Gruhn MA, Dunbar JP, et al. Coping, emotion regulation, and psychopathology in childhood and adolescence: a meta-analysis and narrative review. Psychol Bull 2017 Sep;143(9):939-991 [FREE Full text] [doi: 10.1037/bul0000110] [Medline: 28616996]

79. Domhardt M, Schröder A, Geirhos A, Steubl L, Baumeister H. Efficacy of digital health interventions in youth with chronic medical conditions: a meta-analysis. Internet Interv 2021 Feb 25;24:100373 [FREE Full text] [doi: 10.1016/j.invent.2021.100373] [Medline: 33732626]
80. Reid AE, Carey KB. Interventions to reduce college student drinking: state of the evidence for mechanisms of behavior change. Clin Psychol Rev 2015 Aug;40:213-224 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.006] [Medline: 26164065]
81. Christ C, Schouten MJ, Blankers M, van Schaik DJ, Beekman AT, Wisman MA, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831 [FREE Full text] [doi: 10.2196/17831] [Medline: 32673212]
82. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychother Res 2018 May;28(3):347-355. [doi: 10.1080/10503307.2017.1411628] [Medline: 29224503]
83. Domhardt M, Nowak H, Engler S, Baumel A, Grund S, Mayer A, et al. Therapeutic processes in digital interventions for anxiety: a systematic review and meta-analytic structural equation modeling of randomized controlled trials. Clin Psychol Rev 2021 Sep 13;90:102084. [doi: 10.1016/j.cpr.2021.102084] [Medline: 34610493]
84. 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]
85. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
86. Johansson P, Høglend P, Ulberg R, Amlo S, Marble A, Bøgwald KP, et al. The mediating role of insight for long-term improvements in psychodynamic therapy. J Consult Clin Psychol 2010 Jun;78(3):438-448. [doi: 10.1037/a0019245] [Medline: 20515219]
87. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. BMC Med Res Methodol 2017 Apr 19;17(1):64 [FREE Full text] [doi: 10.1186/s12874-017-0347-z] [Medline: 28420349]
88. Grimm KJ, Ram N. Latent growth and dynamic structural equation models. Annu Rev Clin Psychol 2018 May 07;14:55-89. [doi: 10.1146/annurev-clinpsy-050817-084840] [Medline: 29734829]
89. Hofmann SG, Curtiss JE, Hayes SC. Beyond linear mediation: toward a dynamic network approach to study treatment processes. Clin Psychol Rev 2020 Mar;76:101824 [FREE Full text] [doi: 10.1016/j.cpr.2020.101824] [Medline: 32035297]
90. Straker N. Psychodynamic psychotherapy for cancer patients. J Psychother Pract Res 1997 Nov 25;7(1):1-9 [FREE Full text] [doi: 10.1080/10503309712331331843] [Medline: 9407471]
91. Herrero R, Vara MD, Miragall M, Botella C, García-Palacios A, Riper H, et al. Working Alliance Inventory for Online Interventions-Short Form (WAI-TECH-SF): the role of the therapeutic alliance between patient and online program in therapeutic outcomes. Int J Environ Res Public Health 2020 Aug 25;17(17):6169 [FREE Full text] [doi: 10.3390/ijerph17176169] [Medline: 32854381]
92. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
93. Baumeister H, Montag C. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Switzerland: Springer Nature; 2019.

Abbreviations

CG: control group
DHI: digital health intervention
DHIHP: digital health intervention with a focus on health promotion
DHIPSY: digital health intervention with a psychotherapeutic focus
RCT: randomized controlled trial
RoB: risk of bias

Edited by R Kukafka; submitted 20.04.21; peer-reviewed by J Apolinário-Hagen, M Rassenhofer; comments to author 21.05.21; revised version received 14.07.21; accepted 27.07.21; published 26.11.21.

Please cite as:
Domhardt M, Engler S, Nowak H, Lutsch A, Baumel A, Baumeister H
Mechanisms of Change in Digital Health Interventions for Mental Disorders in Youth: Systematic Review
J Med Internet Res 2021;23(11):e29742
URL: https://www.jmir.org/2021/11/e29742
doi:10.2196/29742
PMID:34842543

Original Paper

The Baby Steps Web Program for the Well-Being of New Parents: Randomized Controlled Trial

David John Kavanagh[1,2], PhD; Jennifer Connolly[1,2], PhD; Jane Fisher[3], PhD; W Kim Halford[4], PhD; Kyra Hamilton[5,6], PhD; Leanne Hides[4], PhD; Jeannette Milgrom[7,8], PhD; Heather Rowe[3], PhD; Paul A Scuffham[6], PhD; Katherine M White[2], PhD; Anja Wittkowski[9], PhD; Shelley Appleton[1,2], PhD; Davina Sanders[1,2], MClinPsych

[1]Centre for Children's Health Research, Queensland University of Technology, South Brisbane Qld, Australia
[2]School of Psychology and Counselling, Queensland University of Technology, South Brisbane Qld, Australia
[3]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[4]School of Psychology, The University of Queensland, Brisbane, Australia
[5]School of Applied Psychology, Griffith University, Brisbane, Australia
[6]Menzies Health Institute Queensland, Griffith University, Brisbane, Australia
[7]Perinatal and Infant Research Institute, Austin Health, Melbourne, Australia
[8]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia
[9]School of Health Sciences, University of Manchester, Manchester, United Kingdom

Corresponding Author:
David John Kavanagh, PhD
Centre for Children's Health Research
Queensland University of Technology
62 Raymond Terrace
South Brisbane Qld, 4101
Australia
Phone: 61 419 479227
Email: david.kavanagh@qut.edu.au

Abstract

Background: New parents face increased risks of emotional distress and relationship dissatisfaction. Digital interventions increase support access, but few preventive programs are optimized for both parents.

Objective: This study aims to conduct the first randomized controlled trial on universal self-guided digital programs to support positive perinatal adjustment of both mothers and fathers. Effects of childcare information (*Baby Care*) and information plus an interactive program (*Baby Steps Wellbeing*) were compared from the third trimester baseline to 3 and 6 months subsequently.

Methods: The study recruited 388 co-parenting male-female adult couples expecting their first single child (26-38 weeks' gestation), using web-based registration. Most (337/388, 86.8%) were obtained from prenatal hospital classes. Couples' randomization was automated and stratified by Edinburgh Postnatal Depression Scale (EPDS) scores (50% couples scored *high* if either mother >7, father >5). All assessments were web-based self-reports: the EPDS and psychosocial quality of life were primary outcomes; relationship satisfaction, social support, and self-efficacy for parenting and support provision were secondary. Linear mixed models provided intention-to-treat analyses, with linear and quadratic effects for time and random intercepts for participants and couples.

Results: Selection criteria were met by 63.9% (248/388) of couples, who were all randomized. Most participants were married (400/496, 80.6%), tertiary educated (324/496, 65.3%), employed full time (407/496, 82%), and born in Australia (337/496, 67.9%). Their mean age was 32.2 years, and average gestation was 30.8 weeks. Using an EPDS cutoff score of 13, 6.9% (18/248) of men, and 16.1% (40/248) of women screened positive for depression at some time during the 6 months. Retention of both partners was 80.6% (201/248) at the 6-month assessments, and satisfaction with both programs was strong (92% ≥50). Only 37.3% (185/496) of participants accessed their program more than once, with higher rates for mothers (133/248, 53.6%) than fathers (52/248, 20.9%; $P<.001$). The EPDS, quality of life, and social support did not show differential improvements between programs, but *Baby Steps Wellbeing* gave a greater linear increase in self-efficacy for support provision ($P=.01$; Cohen $d=0.26$) and lower reduction in relationship satisfaction ($P=.03$; Cohen $d=0.20$) than *Baby Care* alone. Mothers had greater linear benefits in parenting self-efficacy over time than fathers after receiving *Baby Steps Wellbeing* rather than *Baby Care* ($P=.01$; Cohen

d=0.51). However, the inclusion of program type in analyses on parenting self-efficacy and relationship satisfaction did not improve model fit above analyses with only parent gender and time.

Conclusions: Three secondary outcomes showed differential benefits from *Baby Steps Wellbeing*, but for one (parenting self-efficacy), the effect only occurred for mothers, perhaps reflecting their greater program use. Increased engagement will be needed for more definitive testing of the potential benefits of *Baby StepsWellbeing* for perinatal adjustment.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12614001256662; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=367277

(J Med Internet Res 2021;23(11):e23659) doi:10.2196/23659

KEYWORDS

perinatal; depression; prevention; men; self-guided; internet

Introduction

Background

Psychological distress before and after childbirth is common in both parents. Although recent meta-analyses estimate that 12%-17% of recent mothers become depressed [1,2], the rate for fathers is approximately 8% [3]. Depression in one parent poses significant coping challenges for their partner and increases their own risk of depression [4-6]. However, few fathers receive treatment for depression or subclinical distress [7], and research on its treatment is sparse [8,9]. Preventive interventions targeting both parents' distress and promoting quality of life are also likely to benefit the couple's relationship [10,11] as well as their child's emotional and social adjustment [12,13].

Digital delivery may be expected to reduce some of the stigma associated with help seeking, which is especially evident in fathers [7], and allow large-scale access to interventions aimed at preventing or treating distress. Web-based cognitive behavior therapy programs for postpartum depression have demonstrated a moderate average effect size in mothers (Cohen *d*=0.54) [14]. However, a search of published papers identified no research on interactive programs specifically for fathers, although interactive exercises for couples are sometimes included [15]. Even web-based information focusing on new fathers is limited [16,17].

Baby StepsWellbeing is a free, self-guided web program that is optimized for multiple devices and aims to prevent distress and promote the well-being of new parents [18]. It was developed by clinical psychologists and a midwife, with a particular focus on new fathers, and was based on unpublished qualitative research on the issues faced by 10 couples with babies aged 3-6 months and on semistructured interviews with 21 recent parents who reviewed the draft program. To maximize the engagement of new parents, the program contained 4 modules with practical tips on preparing for the baby and addressing issues related to feeding, sleep, and soothing (*Baby Care*). The full *Baby Steps Wellbeing* program had an additional four modules providing information and interactive problem solving on self-care, relationships, interacting with the baby and adjusting to the new roles, as well as a module specifically focused on fathers that covered the ways in which they could support their partner and share childcare tasks. Selecting a tip in *Baby Steps Wellbeing* allowed users to choose specific baby care and well-being tips and set calendar reminders to try them. They could also upload photos about good times with their baby for use at times of stress.

Digital provision of a support program does not guarantee that men will access it, and users of digital mental health services remain predominantly women [19]. Encouraging partners to enroll in a program together may be one way to maximize fathers' engagement. This strategy may also provide opportunities to build mutual support and relationship satisfaction, which are key predictors of perinatal well-being [20]. Accordingly, this study required that both partners participate, protecting their confidentiality through separate usernames and passwords.

Objective

The aim of this randomized controlled trial is to compare the efficacy of the full, interactive *Baby Steps Wellbeing* program (containing all modules and planning tools) with that of *Baby Care* informational tips alone over a period of 6 months from the third trimester of pregnancy onward. The primary outcomes were self-rated depression and psychosocial quality of life, whereas the secondary outcomes included relationship satisfaction, parental self-efficacy, and social support. We predicted that the *Baby Steps Wellbeing* program would result in better outcomes at 3 and 6 months from baseline, compared with *Baby Care* information only. Although we did not have a directional hypothesis about the relative effects of the programs on fathers and mothers, we tested whether the inclusion of an interaction of treatment and parent gender added to the prediction of outcomes. Effects for time were partialed into linear and quadratic effects because of the potential for deterioration in mental health and well-being during the early postpartum period, with some recovery at 6 months. We were interested both in whether the two treatments had differential effects linearly over the study period and whether they modified any tendency for short-term changes at around 2 weeks after childbirth (ie, 3 months from baseline).

Methods

Brief Description

A detailed description of the methodology is provided in the research protocol [21]. A summary is provided below, highlighting the changes from the published protocol.

Ethical Approval, Inclusion Criteria, and Recruitment

The trial received ethical approval from the human research ethics committees of Queensland University of Technology (#1400000687) and Mater Health Services (HREC/14/MHS/166). The participants were co-parenting male-female couples aged 18 years or older who were expecting a single first child at 26-38 weeks' gestation. Recruitment was timed during late pregnancy to engage parents before they faced the heavy demands of childcare [22], while ensuring that their entry was sufficiently proximal to the birth to maximize engagement and later recall. The participants were proficient in English, had access to the internet and a mobile phone, and had completed the baseline assessments. Recruitment occurred between March 2015 and October 2015, and the registration was conducted on the web. The study had a recruitment target of 240 couples, which was expected to enable the detection of a small effect size of f=0.082 (α=.05; power=0.80; repeated measures *r*=0.50, using repeated measures analysis of variance, G*Power 3.1 [Heinrich Heine University] [23]). However, we recruited 248 couples (124 couples in each arm). Most recruited couples (216/248, 87.1%) entered the study through prenatal parent education classes at the Mater Mothers' Hospital, Brisbane, Australia, where a member of the research team presented information about the research trial immediately before the class. The remaining participants were recruited through a trade exhibit at the *Pregnancy, Babies, and Children's Expo* in Brisbane, Australia (25/248, 10.1%); were referred by a friend, family member, or medical practitioner (6/248, 2.4%); or responded to advertising on a parenting website (1/248, 0.4%).

Measures

Demographic and Pregnancy Data

At baseline, each parent reported their age, marital and employment status, education level, country of birth, years living in Australia, and indigenous background (where applicable). The assessment at 3 months after childbirth included weeks of gestation at birth, birth type, baby's sex, and information about any inpatient care of the baby.

Program Use and Satisfaction

The number of log-ins, modules viewed, the number of saved action plans from baseline to 6 months, and the period between the first and last log-in to the program were downloaded from the program's database. Overall program satisfaction and the program's perceived relevance, usefulness, and ease of finding what the participants wanted were self-reported at 3 months on a scale from 0 (not at all) to 100 (extremely), and average scores across these four items are reported below (in this sample, Cronbach α=.93; corrected interitem correlations=0.73-0.91).

Outcomes

Except for the Assessment of Quality of Life-8 Dimensions (AQoL-8D), all self-reported outcomes used unweighted sums or averages of items. The primary outcome measures were the scores derived from the 10-item Edinburgh Postnatal Depression Scale (EPDS) [24,25], which provides a total score of 0-30, with higher scores indicative of depressive symptomatology, and the Psychosocial Super Dimension Scale from the AQoL-8D [26,27], which gives an average weighted score of increasing utility from 0 to 1.

Relationship satisfaction was measured using the 16-item Couples Satisfaction Index (CSI-16) [28], which gives a score of increasing satisfaction from 0 to 81. Social support was assessed using the 4-item version of the Medical Outcomes Study Social Support Survey (Short MOS-SSS) [29], which provides a total score from 4 to 20. A new measure of parenting self-efficacy used an average across the items on feeding, sleep, and settling (eg, "Thinking about the next 13 weeks, how confident are you with putting your baby to sleep?"), each of which was rated on an 11-point scale from 0 (not at all confident) to 100 (extremely confident). At baseline, the scale had high internal consistency (Cronbach α=.94; corrected item-total correlations=0.78-0.95). Self-efficacy for support provision to their partner used a single item ("Thinking about the next 13 weeks, how confident are you with providing support to your partner," which was also rated on an 11-point scale from 0 to 100) that was moderately related to the CSI-16 (Spearman ρ=0.32; $P<.001$) and inversely related to the EPDS (Spearman ρ=–0.30; $P<.001$). Data on income, work time, productivity, and health care were also collected but will be reported in a separate paper.

Web Programs

The participants randomized to the *Baby Care* trial arm received the four informational modules giving information and tips on getting prepared, feeding and soothing their baby, and improving their baby's sleeping habits. A *Get Help* tab provided a list of relevant digital or telephone support services.

The participants receiving *Baby Steps Wellbeing* could also access modules on physical and emotional self-care, their relationship with their partner, changing roles, and interacting with their baby, as well as the module meant especially for fathers (ie, a total of 9 modules, including those focused on baby care). For either *Baby Care* tips or the ones in other modules, the participants in *Baby Steps Wellbeing* could identify goals, solve problems, develop a plan, set times to take action, and record their successful completion. A list of successfully completed plans was available. A web-based scrapbook was used to store photos of good times with their baby, and their dashboard presented due dates for their action plans, together with a rotating quiz question about baby care, a tip, and a scrapbook photo. The database for *Baby Steps Wellbeing* was hosted at the Queensland University of Technology. The participants were not given advice on the number of modules to access, and there was no limitation on the pace of module access. No changes to either program were made during the study, and no major technical issues were encountered.

All participants received automated text messages at 2, 4, 7, and 10 weeks after allocation, reminding them to log in to the program and select tips to apply. Texts sent to the *Baby Steps Wellbeing* participants also included a recommendation to review their goals and plans. A final SMS text message was sent to all participants, thanking them and expressing the hope that they found the website useful.

Procedure

After the volunteers for the study registered on the web, they were emailed a link to the web-based consent form and eligibility screen. Prospective participants could also email or call the research team to ask questions about the trial. After informed consent and initial automated confirmation of eligibility, they were given a link to self-complete the baseline assessments. Both these and later measures were delivered through Qualtrics (Qualtrics, LLC; [30]) and stored separately from identifying data, using a numerical code for the couple plus a letter to signify father or mother. Completion of all items in each measure was required. If a respondent screened positive for a medium or high risk of major depression or self-harm (EPDS item 10 ≥1, or totals of ≥10 for mothers, ≥6 for fathers) at either the baseline or a later assessment, they were contacted by a member of the research team, who implemented a risk management plan [21].

The web-based completion of baseline measures by both the father and mother triggered a fully automated random allocation of the couple to a treatment by Goji, a web-based trial management system developed at Queensland University of Technology [31]. Randomization was performed in permuted blocks, stratified by EPDS (screening negative for anxiety or major or minor depression: mother ≤7 *and* father ≤5; screening positive: mother >7 *or* father >5) [25]. These cutoffs resulted in 123 couples (49.6%) screening positive—ie, approximating a median split of the couples on the EPDS.

At 13 and 26 weeks, the participants were sent automated emails with links to self-complete assessments on the web. All outcome and economic measures were readministered, together with details about the birth, a check on marital status, and ratings of program satisfaction. Nonresponse resulted in emails or calls by a research officer, who was blind to the treatment, to provide further reminders. The participants were given a retail store voucher for Aus $20 (approximately US $16) for each completed follow-up assessment.

Statistical Analyses

The primary analyses adopted an intention-to-treat approach using RStudio version 1.3.1093 lme4 (RStudio, PBC), which provided mixed models analyses of data over the three measurement occasions (baseline, 3 months, and 6 months). Given the potential for correlations within couples, each analysis incorporated random intercepts for both the couple and individual participant. Preliminary analyses demonstrated that the inclusion of a first-order autocorrelation between test occasions did not improve predictions; therefore, it was not included in the reported analyses. As mothers are somewhat more vulnerable than fathers to issues with perinatal mental health and well-being [1-3], all analyses included effects for parent gender and the interaction of parent gender and time. In compliance with CONSORT (Consolidated Standards of Reporting Trials) guidelines, analyses of outcomes did not control for any differences between the treatments at baseline (contrary to the protocol paper) [21], but tests for any baseline differences are reported in the interests of full disclosure.

In the primary analyses, we compared two models: (1) a hypothesized model, examining the relative impact of the two programs, controlling for parent gender (ie, including effects of treatment×time, parent gender×time, and the relative main effects) and (2) the full factorial model, testing all potential effects. Post hoc analyses were also undertaken to examine a model that only included parent gender and time. The reported *t* tests used the Satterthwaite method, and the degrees of freedom were calculated using the Kenward-Roger method. The effect sizes in these analyses use Cohen *d*, with difference scores being divided by shared baseline SD units (taken from analysis of variance values of baseline scores, with treatment and parent gender as independent variables).

Results

Participant Characteristics

The CONSORT diagram is presented in Figure 1. Of the 726 people who expressed initial interest in the study, 388 couples registered; 248 (53.4) fulfilled the eligibility criteria and were randomly allocated to the two treatments. As shown in Figure 1, retention at 3- and 6-month assessments was high in both treatments. One couple had to be withdrawn from the study because of a stillbirth, but their baseline data were still included in the intention-to-treat analyses. An attempt was made to contact participants who screened positive for risk of depression or self-harm and implement the risk management procedure (112 at baseline, 175 at 3 months, 132 at 6 months), but no one had to be withdrawn from the study for this reason. Very few participants received concurrent mental health treatment during the study in either treatment (medication: 5%-7% and counseling: 4%-6%).

On average, the participants receiving *Baby Steps Wellbeing* entered the study a little less than a week earlier in the pregnancy compared with those receiving *Baby Care* (mean 30.4, SD 2.8 vs *Baby Care* mean 31.2, SD 3.1; $F_{1,246}$=4.35; *P*=.04; η^2=0.017), but there were no other significant differences between the treatments in terms of demographic characteristics. Tests of any differences between the fathers and mothers in the study are presented in Tables 1 and 2. Mothers were slightly younger than fathers on average and were more likely to have a university degree but less likely to be employed full time. At baseline, they had higher EPDS scores, lower AQoL-8D Psychosocial Super Dimension scores, and lower self-efficacy in the provision of social support. However, they had higher relationship satisfaction scores on the CSI-16 and higher perceived social support on the Short MOS-SSS.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.

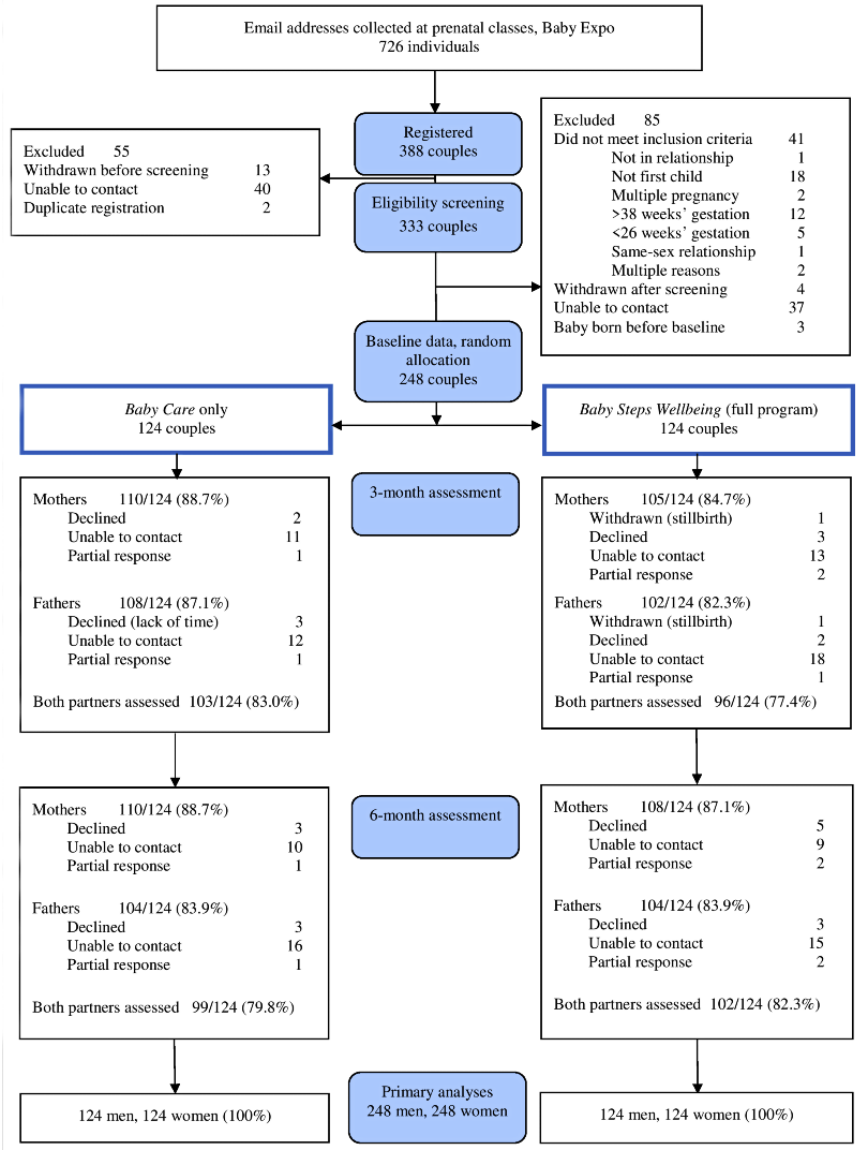


Table 1. Baseline characteristics and scores by parent gender (fathers: n=248; mothers: n=248): categorical variables.

Variable	Fathers, n (%)	Mothers, n (%)	Chi-square test (*df*)	*P* value
University degree	142 (57.2)	182 (73.3)	14.24 (1)	<.001
Employment			13.98 (2)	.001
Full time	219 (88.3)	188 (75.8)		
Part time or casual	20 (8.0)	35 (14.1)		
Unpaid leave, benefits or retired	9 (3.6)	25 (10.0)		
Indigenous	4 (1.6)	0 (0)	4.03 (1)	.045
Region of birth			4.12 (2)	.13
Australia	163 (65.7)	174 (70.2)		
New Zealand or Pacific or United Kingdom or Ireland or North America or other European countries	53 (21.4)	36 (14.5)		
Asia or Southeast Asia or Middle East or South America or Africa	32 (12.9)	38 (15.3)		

Table 2. Baseline characteristics and scores by parent gender (fathers: n=248; mothers: n=248): continuous variables.

Variable[a]	Fathers, predicted mean (SE)	Mothers, predicted mean (SE)	*t* test (*df*)	*P* value
Age (years)	33.0 (0.27)[b]	31.5 (0.27)[b]	15.77 (247.0)	<.001
Years living in Australia[c]	12.2 (1.02)[d]	10.5 (1.08)[e]	–1.37 (68.6)	.18
Depression (EPDS[f] total)	4.47 (0.25)	5.42 (0.25)	2.81 (247.0)	.005
Quality of life (AQoL-8D[g] Psychosocial)	0.474 (0.01)	0.434 (0.10)	–3.13 (247.0)	.002
Relationship satisfaction (CSI-16[h])	69.8 (0.60)	71.9 (0.60)	3.26 (247.0)	.001
Social support (Short MOS-SSS[i])	15.8 (0.21)	16.6 (0.21)	2.81 (247.0)	.005
Parenting self-efficacy	65.2 (1.20)	65.3 (1.20)	0.08 (247.0)	.94
Self-efficacy: support provision	81.2 (1.26)	71.8 (1.26)	–5.43 (247.0)	<.001

[a]Analyses of continuous variables include a random intercept for a couple.

[b]N=248.

[c]If born overseas.

[d]n=85.

[e]n=74.

[f]EPDS: Edinburgh Postnatal Depression Scale.

[g]AQoL-8D: Assessment of Quality of Life-8 Dimensions.

[h]CSI-16: Couples Satisfaction Index-16.

[i]MOS-SSS: Medical Outcomes Study Social Support Survey.

The participants had a median age of 32 (mean 32.2, SD 4.4; range 20-47) years, and 80.6% (400/496) were married. Most had a university degree (324/496, 65.3%), were employed full time (407/496, 82%), had been born in Australia (337/496, 67.9%), or had lived in Australia for a substantial period (mean 11.2 years, SD 9.5 years). Using a cutoff score of 12/13 on the EPDS [5] and no data substitution, 16.1% (40/248) of the mothers and 6.9% (18/248) of the fathers screened positive for depression at some time during the study. The rates at 3 months (approximately 2 weeks after childbirth) were 9.1% (20/217) for the mothers and 2.9% (6/209) for the fathers. If the Australian cutoffs for major or minor depression reported in the study by Matthey et al [25] were used (8/9 for mothers and 9/10 for fathers), 41.5% (103/248) of the mothers and 18.5% (46/248) of the fathers screened positive at some time, with the rates at 3 months being 27.2% (59/217) and 9.1% (19/209), respectively. The details are presented in Multimedia Appendix 1, Table S1.

Program Satisfaction and Use

Satisfaction with the program was high among the participants who accessed it at least once (median 75, 92% ≥50; Table 3), with no substantial differences due to treatment or parent gender, but access and use were suboptimal, especially for the fathers. The median number of log-ins across the sample was 1 (range 0-23; 90th percentile=5). Only 36.7% (91/248) of *Baby Care* and 37.9% (94/248) of *Baby Steps Wellbeing* participants (χ^2_1=0.08; *P*=.781) accessed programs on two or more occasions. While 53.6% (133/248) of the mothers used a program at least twice, only 21.0% (52/248) of the fathers did so (χ^2_1=56.56; *P*<.001). Among the *Baby Steps Wellbeing* participants, 44.4% (55/124) of the mothers set a goal, whereas only 12.9% (16/124) of the fathers did so (χ^2_1=30.02; *P*<.001). These parent gender differences were observed for all continuous variables related to program use (Table 3; Multimedia Appendix 1, Table S2). Despite the fact that the fathers receiving *Baby Steps Wellbeing* could access more than twice the number of modules (nine) than those receiving *Baby Care* (four), fathers only accessed an average of just one module in either treatment. The significant interaction between treatment and parent gender (t_{246}=2.35; *P*=.020) was a reflection of the fact that the mothers took greater advantage of the additional modules, although on average they still accessed fewer than half of those in *Baby Steps Wellbeing* (Table 2). More *Baby Care* modules were viewed by the participants in the *Baby Care* treatment (t_{492}=–3.40; *P*<.001), but there were no statistically significant differences between the treatments in terms of log-ins or duration of use (Multimedia Appendix 1, Table S2). The baseline EPDS scores had a negligible association with the use of the programs (Multimedia Appendix 1, Table S2), but a weak negative correlation with program satisfaction (*r*=–0.18; *P*=.03).

Table 3. Program use and satisfaction[a].

Variable	Values, predicted mean (SE)			
	Baby Care only		Baby Steps Wellbeing	
	Fathers	Mothers	Fathers	Mothers
Log-ins	1.09 (0.24)	2.67 (0.24)	0.89 (0.24)	2.77 (0.24)
Duration of use (days)	10.87 (4.08)	34.62 (4.08)	12.27 (4.08)	41.47 (4.08)
Modules viewed				
Total	0.98 (0.17)	2.23 (0.17)	1.07 (0.17)	3.10 (0.17)
Baby Care	0.98 (0.12)	2.23 (0.12)	0.40 (0.12)	1.56 (0.12)
Program satisfaction[b]	69.2 (2.03)	71.8 (2.01)	70.8 (2.10)	72.5 (2.06)

[a]Predicted means are derived from analyses that include a random intercept for a couple.

[b]Mean of responses across participants who had accessed a program to items on overall satisfaction, relevance, usefulness, and ease of finding what they wanted.

Model Comparisons

Initially, a comparison was made of model fit from the hypothesized model and from a full factorial model. With the exception of parenting self-efficacy, the hypothesized model gave a better fit than the full factorial model in terms of both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and the simpler model did not result in a significantly poorer likelihood ratio (Table 4). In the case of parenting self-efficacy, AIC favored the full factorial model. Although penalization for the greater number of predictors resulted in a slightly poorer BIC value, a better fit from the full model approached significance on the likelihood ratio test (P=.05). Accordingly, the subsequent main analyses on this variable used the full factorial model, whereas the hypothesized model was used for the remaining outcome variables. Results for each variable using the nonpreferred model are displayed in Multimedia Appendix 1, Table S3.

In additional post hoc analyses, we also tested the fit from a model with only parent gender and time to assess whether any significant effects for treatment resulted in an improved model fit overall (Multimedia Appendix 1, Table S4). For self-efficacy for support provision, this model gave a significantly poorer fit and a larger AIC value than the hypothesized model, although the reduction in model complexity resulted in a slightly lower BIC value. For the remaining outcomes, the AIC and BIC values were slightly smaller for the simpler model, but the likelihood ratio test was not significantly different. We note that these results for the model fit moderate the conclusions from the analyses on the effects of treatment on relationship satisfaction (CSI-16) and parenting self-efficacy, which are described below.

Table 4. Comparison of hypothesized and full factorial models for analyses of outcomes[a].

Dependent variable	Parameters, N	AIC[b]	BIC[c]	Log likelihood	Deviance	Chi-square test (*df*=3)	*P* value
Depression (EPDS[d])							
Hypothesized model[e]	12	7247.7	7310.2	–3611.9	7223.7	3.34	.34
Full model	15	7250.4	7328.5	–3610.2	7220.4	3.34	.34
Quality of life (AQoL-8D[f] Psychosocial)							
Hypothesized model[e]	12	–1361.4	–1298.9	692.7	–1385.4	4.90	.18
Full model	15	–1360.3	–1282.2	695.2	–1390.3	4.90	.18
Relationship satisfaction (CSI-16[g])							
Hypothesized model[e]	12	9716.4	9778.9	–4846.2	9692.4	3.94	.27
Full model	15	9728.5	9796.6	–4844.2	9688.5	3.94	.27
Social support (Short MOS-SSS[h])							
Hypothesized model[e]	12	6826.2	6888.7	–3401.1	6802.2	2.79	.42
Full model	15	6829.4	6907.5	–3399.7	6799.4	2.79	.42
Parenting self-efficacy							
Hypothesized model	12	11395	11458	–5685.7	11371	7.65	.05
Full model[e]	15	11394	11472	–5681.9	11364	7.65	.05
Self-efficacy for support provision							
Hypothesized model[e]	12	11560	11622	–5768.0	11536	4.03	.26
Full model	15	11562	11640	–5765.9	11532	4.03	.26

[a]All models had random intercepts for the subject and couple. The hypothesized model included treatment, parent gender, time, treatment×time, and parent gender×time. The full model comprised the full factorial design.

[b]AIC: Akaike Information Criterion.

[c]BIC: Bayesian Information Criterion.

[d]EPDS: Edinburgh Postnatal Depression Scale.

[e]Preferred model.

[f]AQoL-8D: Assessment of Quality of Life-8 Dimensions.

[g]CSI-16: Couples Satisfaction Index-16.

[h]MOS-SSS: Medical Outcomes Study Social Support Survey.

Primary Outcomes

The effects for each outcome variable from the preferred model are displayed in Table 5, and the means are shown in Figure 2 and Multimedia Appendix 1, Table S5. Over the course of the study, the mothers had higher EPDS scores for depression (Cohen *d*=0.37) and a lower average quality of life on the AQoL-8D Psychosocial Super Dimension than the fathers (Cohen *d*=0.27). The EPDS scores did not significantly change across the sample from baseline to 6 months, but the mothers' depression had a greater tendency to peak soon after childbirth, as shown by a significant parent gender×quadratic time effect (Table 5). For example, from baseline to 3 months, their EPDS scores rose 0.27 SD units more than those of the fathers. There was a linear improvement in quality of life over the study period (Cohen *d*=0.27), which once again was modified by an interaction between parent gender and quadratic time (eg, from baseline to 3 months): the psychosocial quality of life of the fathers increased by 0.16 SD units more than that of the mothers. However, there were no significant interactions between treatment and time on either primary outcome when the hypothesized model was used.

Table 5. Outcomes, using the preferred model.

Dependent variable	Estimate	SE	*t (df)*	*P* value
Depression (EPDS[a])				
Intercept	0.6744	0.2431	–2.77 (427.0)	.006
Treatment	–0.1820	0.2875	–0.63 (247.2)	.53
Parent gender	1.4675	0.2656	5.53 (246.6)	<.001
Linear time	–0.2993	0.2384	–1.26 (899.5)	.21
Quadratic time	0.0390	0.2417	0.16 (883.2)	.87
Parent gender×linear time	0.3537	0.2735	1.29 (896.2)	.196
Parent gender×quadratic time	–0.6512	0.2799	–2.33 (886.6)	.02
Treatment×linear time	–0.2681	0.2737	–0.98 (894.7)	.33
Treatment×quadratic time	–0.2556	0.2800	–0.91 (885.4)	.36
Psychosocial quality of life (AQoL-8D[b] Psychosocial)				
Intercept	0.0142	0.0120	1.18 (381.2)	.24
Treatment	0.0225	0.0145	1.50 (247.7)	.13
Parent gender	–0.0445	0.0114	–3.89 (249.3)	<.001
Linear time	0.0282	0.0091	3.11 (888.9)	.002
Quadratic time	–0.0073	0.0092	–0.79 (876.9)	.43
Parent gender×linear time	0.0088	0.0104	0.85 (888.7)	.398
Parent gender×quadratic time	0.0265	0.0107	2.48 (880.2)	.01
Treatment×linear time	–0.0038	0.0104	–0.37 (884.3)	.71
Treatment×quadratic time	0.0090	0.0107	0.84 (877.1)	.40
Relationship satisfaction (CSI-16[c])				
Intercept	–0.8547	0.8088	–1.06 (336.2)	.29
Treatment	0.5978	1.0520	0.57 (247.3)	.57
Parent gender	1.0437	0.6333	1.65 (244.8)	.101
Linear time	–2.3895	0.5498	–4.35 (887.7)	<.001
Quadratic time	0.1802	0.5564	0.32 (874.1)	.75
Parent gender×linear time	–1.3776	0.6293	–2.19 (890.0)	.03
Parent gender×quadratic time	0.2033	0.6445	0.32 (880.1)	.75
Treatment×linear time	1.3444	0.6323	2.13 (880.7)	.03
Treatment×quadratic time	–0.4098	0.6465	–0.63 (873.4)	.53
Social support (Short MOS-SSS[d])				
Intercept	–0.5233	0.2177	–2.40 (372.9)	.02
Treatment	0.3073	0.2681	1.15 (235.3)	.25
Parent gender	0.7202	0.2135	3.37 (243.9)	<.001
Linear time	–0.2292	0.2044	–1.12 (882.1)	.26
Quadratic time	0.17140	0.2072	0.83 (865.8)	.41
Parent gender×linear time	–0.1775	0.2344	–0.76 (880.1)	.45
Parent gender×quadratic time	–0.1654	0.2402	–0.69 (870.6)	.49
Treatment×linear time	0.1808	0.2350	0.77 (876.9)	.44
Treatment×quadratic time	–0.0543	0.2407	–0.23 (867.7)	.82
Parenting self-efficacy[e]				

Dependent variable	Estimate	SE	*t (df)*	*P* value
Intercept	–0.9291	1.1960	–0.78 (472.2)	.44
Treatment	–0.2268	1.6960	–0.13 (477.2)	.89
Parent gender	3.2155	1.6180	1.99 (238.7)	.048
Linear time	11.3882	1.3400	8.50 (889.7)	<.001
Quadratic time	–1.3224	1.3535	–0.98 (868.3)	.33
Parent gender×Treatment	0.0261	2.2939	0.01 (240.8)	.99
Parent gender×linear time	1.9064	1.8762	1.02 (883.0)	.31
Parent gender×quadratic time	0.1665	1.9024	0.09 (867.1)	.93
Treatment×linear time	–3.3766	1.8945	–1.78 (889.1)	.08
Treatment×quadratic time	–1.7702	1.9429	–0.91 (878.4)	.36
Treatment×parent gender×linear time	6.8500	2.6569	2.58 (883.5)	.01
Treatment×parent gender×quadratic time	2.9442	2.7200	1.08 (874.5)	.28
Self-efficacy for support provision				
Intercept	3.8583	1.1927	3.24 (418.2)	.001
Treatment	–0.7276	1.4183	–0.51 (245.0)	.61
Parent gender	–6.7124	1.2864	–5.22 (241.3)	<.001
Linear time	–0.5999	1.1922	–0.50 (893.7)	.62
Quadratic time	1.4951	1.2085	1.24 (877.1)	.22
Parent gender×linear time	3.6372	1.3681	2.66 (890.8)	.008
Parent gender×quadratic time	–0.2716	1.3996	–0.19 (880.9)	.85
Treatment×linear time	3.5459	1.3695	2.59 (888.7)	.01
Treatment×quadratic time	–2.2494	1.4006	–1.61 (879.4)	.11

[a]EPDS: Edinburgh Postnatal Depression Scale.

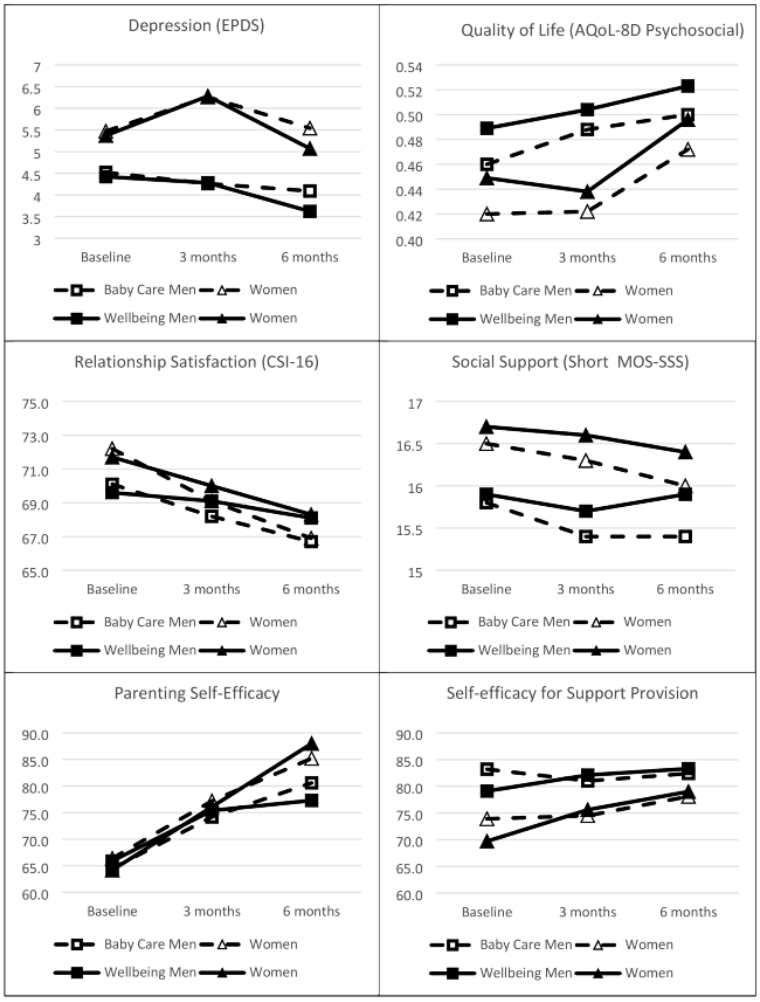
[b]AQoL-8D: Assessment of Quality of Life-8 Dimensions.

[c]CSI-16: Couples Satisfaction Index-16.

[d]MOS-SSS: Medical Outcomes Study Social Support Survey.

[e]The reported effects are from the hypothesized model, except for parenting self-efficacy, which is from the full factorial model.

Figure 2. Predicted mean outcomes, using the preferred model. Estimations of means used the hypothesized model, except for parenting self-efficacy, which used the full factorial model. AQoL-8D: Assessment of Quality of Life-8 Dimensions; CSI-16: Couples Satisfaction Index-16; EPDS: Edinburgh Postnatal Depression Scale; MOS-SSS: Medical Outcomes Study Social Support Survey.



Secondary Outcomes

Relationship satisfaction on the CSI-16 fell over the course of the study (Cohen *d*=–0.36), especially for the mothers (Cohen *d*=–0.46 vs –0.26 for the fathers). However, the reduction in relationship satisfaction was less in *Baby Steps Wellbeing* than in *Baby Care* (Cohen *d*=0.20). Significant interactions with treatment were also observed in the two self-efficacy measures. In the case of self-efficacy for support provision, the *Baby Steps Wellbeing* participants showed a greater linear increase over time than those in *Baby Care* (Cohen *d*=0.26). Across the whole study, the mothers had lower scores than the fathers (Cohen *d*=–0.33), but their scores improved to a greater extent over the study period (Cohen *d*=0.34 vs 0.08 for the fathers). The interaction of time and parent gender was not further modified by treatment. In contrast, the mothers had higher scores on parenting self-efficacy than the fathers on average (Cohen *d*=0.17), as well as showing a greater linear improvement over time (Cohen *d*=1.12 vs 0.72 for the fathers). In this case, there was also a three-way interaction: as presented in Figure 2, the mothers tended to have greater increases in *Baby Steps Wellbeing* than in *Baby Care* (Cohen *d*=0.26), but the reverse was the case for the fathers (Cohen *d*=–0.25). The only secondary outcome not showing any interactions between treatment and time was social support on the Short MOS-SSS. The mothers scored somewhat more highly than the fathers across the study period (Cohen *d*=0.22), but there were no other significant effects.

Discussion

Principal Findings

Although satisfaction with the *Baby Care* and *Baby Steps Wellbeing* programs was moderately high, the average number of log-ins and viewed modules was suboptimal, especially for the fathers. Joint access to the program by some partners may have led to underestimates of individual access, but little use was made of the key planning and goal-setting segments of *Baby Steps Wellbeing*.

Consistent with this limited access, perinatal distress on the EPDS, psychosocial quality of life, and social support on the Short MOS-SSS did not show significant differential improvements for the two treatments. However, there was some evidence of differential benefits from the full, interactive *Baby Steps Wellbeing* program compared with the *Baby Care* information modules in 3 of the 4 secondary outcomes. The participants who received the full program had a greater linear rise in self-efficacy for support provision and an attenuated

reduction in relationship satisfaction over the study than those who only received *Baby Care* information. Furthermore, consistent with greater program access by the mothers, they had greater linear increases than the fathers in their parenting self-efficacy if they received *Baby Steps Wellbeing* than if they were only given *Baby Care* information. These results are partially consistent with those from a recent universal prevention trial on parental stress, which used a booklet, video, home visit, and telephone call focused on baby care and being sensitive to the needs of both the parents and the baby and did not detect benefits over usual care on the EPDS [32]. However, that earlier study did not find differential effects on any other outcomes, including a single-item measure of parenting self-efficacy.

The interpretation of greater benefits from the full *Baby Steps Wellbeing* program is moderated by small effect sizes and by the fact that self-efficacy for support provision comprised a new single-item measure. In addition, the results from analyses in Multimedia Appendix 1, Table S3 that used a nonpreferred model differ from those in the preferred-model analyses, casting doubt on the robustness of the reported effects, and post hoc analyses showed that the significant interactions between time and treatment for 2 outcomes (relationship satisfaction and parenting self-efficacy) were insufficient to give a better model fit compared with a model that only included parent gender and time. Furthermore, all significant interactions between treatment and time involved linear changes to 6 months, rather than specifically modifying outcomes immediately after childbirth, when the greatest vulnerability is seen.

However, the fact that some significant effects were seen at all remained noteworthy, given the limited degree of program involvement and the fact that this was an unselected sample, with average scores on all outcome variables showing little dysfunction throughout the study. In this context, the lack of significant differential effects of the treatments on EPDS scores should not have been surprising, given the study's low point-prevalence of screening positive for self-reported depression (3% of the fathers and 9% of the mothers at 3 months, using an EPDS cutoff at 12/13) compared with the rates in meta-analyses (8% of fathers [3] and 12%-17% of mothers [1,2]), which restricted the opportunity to detect differential changes.

There were insufficient couples who both accessed and set goals for multiple *Baby StepsWellbeing* modules or continued to use the program postnatally (ie, for more than approximately 10 weeks) to allow a post hoc examination of whether more sustained or intensive program use would have provided stronger results, but this study's results clearly suggest that these should be goals for future research. Our program had already attempted to maximize ongoing program engagement in multiple ways: (1) by recruiting couples who might reinforce each other's engagement in program use and resultant actions; (2) by providing information related to baby care, which we expected to have salience late in the pregnancy; (3) by repeatedly reminding participants to log in and giving brief tips that changed at each visit; and (4) by offering opportunities to upload photos that would cue memories of good times with their baby and make it pleasurable to visit the site. We timed entry into the study and initial program exposure while the women were pregnant because of the time pressures experienced after childbirth, but solutions to childcare or general well-being issues might not have immediately seemed relevant to an unselected sample at that time. Furthermore, recruitment primarily through prenatal classes could have resulted in redundancy in some baby care information. In fact, a median of 1-2 website visits suggested that many participants may have regarded the program as just one of many web-based information repositories on baby care, which they only used once.

The absence of a predetermined, sequential progression through the program's modules might have been a significant factor in the limited use of the program [33,34]. Gamification of progression may increase program use (eg, rewarding module completion or progressive gains in knowledge or skills and inclusion of animations or videos or audios that modeled functional problem solutions [35,36]). The positive effects of coaching on maintained program engagement [17] could be simulated by digital *coaches* who might offer some features that approximated human interaction (eg, greeting and praising completion of tasks) [37], even if the individualization of these responses was limited. Although motivational interviewing before beginning the program may be more feasible in a depression treatment program than in a program that seeks to prevent depression [35], a brief digital adaptation of motivational interviewing (together with a rationale for regular program use) may also assist, especially in relation to addressing the couple's well-being. Impact and perceived relevance of the program might also be increased if the SMS text messages incorporated tips [38] or empowering messages rather than just reminding users to return to the program.

Problems with engaging fathers in prevention or treatment of mental health conditions are endemic, both perinatally and more generally [8,9], and occur for digital interventions as well as face-to-face services [19]. We had hoped that the provision of a module especially for fathers would increase its perceived relevance and that parallel use of the program by mothers would encourage its use by fathers. However, it seems that further refinement of the approach with a more extensive co-design is required.

There were several indications that the mothers reported more pronounced mental health and well-being issues than the fathers in this study, particularly in the perinatal period. They had higher average EPDS scores as well as lower psychosocial quality of life and self-efficacy for support provision over the course of the study, together with a greater tendency for their EPDS scores to peak. Their quality of life rose less strongly soon after childbirth, and their relationship satisfaction fell over time. Although they had slightly higher average social support over the course of the study and (not unexpectedly) were more confident about their parenting skills than the fathers, these gender differences further substantiate the particular vulnerability of new mothers to issues with mental health and well-being [1-3].

Strengths and Limitations

To our knowledge, this is the first controlled trial that specifically attempted to engage fathers in a digital perinatal intervention and provided an interactive module especially for

them. The trial had a substantial sample size and a very high retention in postbaseline assessments. Randomization was performed through an automated trial management program that ensured equivalence between treatment groups of the EPDS scores at baseline. All assessments were conducted on the web, and the follow-up research staff were blinded to the participants' allocation. We conducted intention-to-treat analyses, with secondary analyses examining effects in the participants who accessed the program. However, the study's predominant recruitment through prenatal classes and the sample's high rate of university education mean that the results may not be generalizable to populations that are less highly educated or less engaged in learning about birthing and childcare. Although recruitment during the third trimester was intended to maximize the perceived relevance of the content, this may also have reduced its perceived novelty as well as excluding parents with early births or pregnancy complications from the sample. The exclusion of parents who were not cohabiting, were not in a male-female relationship, were aged below 18 years, or were expecting multiple infants also restricted the potential application of this study.

Conclusions

There was some evidence of differential benefits from the full *Baby Steps Wellbeing* program, although these benefits were more consistently seen in the mothers than in the fathers, and the effect sizes were relatively small. A major contributor to these results may have been the low level of use of the self-guided program, especially by the fathers. Increasing this program's use will be a prerequisite for more accurate future estimation of the degree of potential impact of the *Baby Steps Wellbeing* program on the well-being of new parents.

Acknowledgments

This trial was funded by the men's health charity Movember through *Beyond Blue* (project grant: KAVA11NPD). The sponsor had no role in the development of the program; in the design, conduct, or analysis of the trial; or in preparing this manuscript for publication. The authors acknowledge the contributions of Anne Chang, Lynne Daniels, and Jennifer Connolly to the study proposal; the contributions of Leigh Davis, Amy Kelly, Elizabeth Hurrion, and Jane Swan to program content; the programming of Steven Edge and Jeremy Gibson; and the graphic design of Brendan Pearce. The program development was informed by the prior qualitative research of Lindsay Bennett. Trial recruitment was assisted by Anne Tremellen, Patricia Wilson, Katy Woodcock, and midwives from the Mater Mothers' Hospital. The authors also acknowledge the contribution of 41 parents who informed the program's content and design.

Authors' Contributions

All authors contributed to the conceptualization and design of the web program and trial, to the conduct of the trial, and the paper. The research officers for this trial were DS, SA, Georgia Stein, Elloise Brake and David Rodwell. Analyses were undertaken by JM and DJK, with advice from Dimitrios Vagenas and Shu-Kay Angus Ng. Joshua Byrnes and Peter Lee undertook cost-effectiveness analyses that will be reported in a separate paper.

Conflicts of Interest

The authors declare no competing financial interests. DJK, JF, JM, and HR collaborate in the development of digital programs and apps to treat or prevent perinatal depression or distress. AW is a member of the Triple P Research Network and has advised the Triple P International Scientific Advisory Committee in a research capacity. LH is supported by a senior research fellowship from the National Health and Medical Research Council (APP1119098).

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 70 KB - jmir_v23i11e23659_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 870 KB - jmir_v23i11e23659_app2.pdf]

References

1. Shorey S, Chee CY, Ng ED, Chan YH, Tam WW, Chong YS. Prevalence and incidence of postpartum depression among healthy mothers: a systematic review and meta-analysis. J Psychiatr Res 2018 Sep;104:235-248. [doi: 10.1016/j.jpsychires.2018.08.001] [Medline: 30114665]
2. Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. J Affect Disord 2017 Sep;219:86-92. [doi: 10.1016/j.jad.2017.05.003] [Medline: 28531848]
3. Cameron EE, Sedov ID, Tomfohr-Madsen LM. Prevalence of paternal depression in pregnancy and the postpartum: an updated meta-analysis. J Affect Disord 2016 Dec;206:189-203. [doi: 10.1016/j.jad.2016.07.044] [Medline: 27475890]

4. Da Costa D, Danieli C, Abrahamowicz M, Dasgupta K, Sewitch M, Lowensteyn I, et al. A prospective study of postnatal depressive symptoms and associated risk factors in first-time fathers. J Affect Disord 2019 Apr 15;249:371-377. [doi: 10.1016/j.jad.2019.02.033] [Medline: 30818245]
5. Edward K, Castle D, Mills C, Davis L, Casey J. An integrative review of paternal depression. Am J Mens Health 2015 Jan;9(1):26-34 [FREE Full text] [doi: 10.1177/1557988314526614] [Medline: 24626601]
6. Ruffell B, Smith DM, Wittkowski A. The experiences of male partners of women with postnatal mental health problems: a systematic review and thematic synthesis. J Child Fam Stud 2019 Jul 13;28(10):2772-2790. [doi: 10.1007/s10826-019-01496-4]
7. Thorsteinsson EB, Loi NM, Farr K. Changes in stigma and help-seeking in relation to postpartum depression: non-clinical parenting intervention sample. PeerJ 2018 Nov 8;6:e5893 [FREE Full text] [doi: 10.7717/peerj.5893] [Medline: 30425892]
8. O'Brien AP, McNeil KA, Fletcher R, Conrad A, Wilson AJ, Jones D, et al. New fathers' perinatal depression and anxiety-treatment options: an integrative review. Am J Mens Health 2017 Jul;11(4):863-876 [FREE Full text] [doi: 10.1177/1557988316669047] [Medline: 27694550]
9. Philpott LF, Savage E, Leahy-Warren P, FitzGearld S. Paternal perinatal depression: a narrative review. Int J Mens Soc Community Health 2020 Mar 06;3(1):e1-15. [doi: 10.22374/ijmsch.v3i1.22]
10. Nishimura A, Fujita Y, Katsuta M, Ishihara A, Ohashi K. Paternal postnatal depression in Japan: an investigation of correlated factors including relationship with a partner. BMC Pregnancy Childbirth 2015 May 31;15:128 [FREE Full text] [doi: 10.1186/s12884-015-0552-x] [Medline: 26026489]
11. Wee KY, Skouteris H, Pier C, Richardson B, Milgrom J. Correlates of ante- and postnatal depression in fathers: a systematic review. J Affect Disord 2011 May;130(3):358-377. [doi: 10.1016/j.jad.2010.06.019] [Medline: 20599275]
12. Gutierrez-Galve L, Stein A, Hanington L, Heron J, Ramchandani P. Paternal depression in the postnatal period and child development: mediators and moderators. Pediatrics 2015 Feb;135(2):e339-e347. [doi: 10.1542/peds.2014-2411] [Medline: 25560437]
13. Hickey G, McGilloway S, Leckey Y, Furlong M, Leavy S, Stokes A, et al. Mothers' well-being, parenting attitudes, and home environment: cumulative risk and parity in early motherhood. Child Care Health Dev 2019 Jul;45(4):523-530. [doi: 10.1111/cch.12677] [Medline: 31026334]
14. Roman M, Constantin T, Bostan CM. The efficiency of online cognitive-behavioral therapy for postpartum depressive symptomatology: a systematic review and meta-analysis. Women Health 2020 Jan;60(1):99-112. [doi: 10.1080/03630242.2019.1610824] [Medline: 31057080]
15. Milgrom J, Schembri C, Ericksen J, Ross J, Gemmill AW. Towards parenthood: an antenatal intervention to reduce depression, anxiety and parenting difficulties. J Affect Disord 2011 May;130(3):385-394. [doi: 10.1016/j.jad.2010.10.045] [Medline: 21112641]
16. Fletcher R. Field testing of father-inclusive guidelines for web-based information and support aimed at families with perinatal depression. Health Promot J Austr 2011 Dec;22(3):231-233. [doi: 10.1071/he11231] [Medline: 22497069]
17. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
18. Kavanagh D, Connolly J, Edge S, Gibson J, Pearce B, Sanders D. The Baby Steps program for new parents. Baby Steps. URL: https://www.babysteps.org.au [accessed 2021-08-19]
19. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
20. Pilkington PD, Milne LC, Cairns KE, Lewis J, Whelan TA. Modifiable partner factors associated with perinatal depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2015 Jun 01;178:165-180. [doi: 10.1016/j.jad.2015.02.023] [Medline: 25837550]
21. Hamilton K, Kavanagh D, Connolly J, Davis L, Fisher J, Halford K, et al. Baby steps - an online program promoting the well-being of new mothers and fathers: a study protocol. JMIR Res Protoc 2016 Jul 01;5(3):e140 [FREE Full text] [doi: 10.2196/resprot.5706] [Medline: 27370711]
22. Goodman JH. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. Birth 2009 Mar;36(1):60-69. [doi: 10.1111/j.1523-536X.2008.00296.x] [Medline: 19278385]
23. 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]
24. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
25. Matthey S, Barnett B, Kavanagh DJ, Howie P. Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. J Affect Disord 2001 May;64(2-3):175-184. [doi: 10.1016/s0165-0327(00)00236-6] [Medline: 11313084]
26. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient 2014;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]

27. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. Qual Life Res 2014 Oct;23(8):2395-2404. [doi: 10.1007/s11136-014-0686-8] [Medline: 24719017]
28. Funk JL, Rogge RD. Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. J Fam Psychol 2007 Dec;21(4):572-583. [doi: 10.1037/0893-3200.21.4.572] [Medline: 18179329]
29. Gjesfjeld CD, Greeno CG, Kim KH. A confirmatory factor analysis of an abbreviated social support instrument: the MOS-SSS. Res Social Work Pract 2007 Nov 29;18(3):231-237. [doi: 10.1177/1049731507309830]
30. Experience design experience improvement. Qualtrics. URL: https://www.qualtrics.com/au/?rid=ip&prevsite=en&newsite=au&geo=IN&geomatch=au [accessed 2021-08-19]
31. The Goji trial management system. Goji. 2014. URL: http://goji.qut.edu.au/ [accessed 2021-08-19]
32. Missler M, van Straten A, Denissen J, Donker T, Beijers R. Effectiveness of a psycho-educational intervention for expecting parents to prevent postpartum parenting stress, depression and anxiety: a randomized controlled trial. BMC Pregnancy Childbirth 2020 Oct 31;20(1):658 [FREE Full text] [doi: 10.1186/s12884-020-03341-9] [Medline: 33129314]
33. Drozd F, Haga SM, Brendryen H, Slinning K. An internet-based intervention (Mamma Mia) for postpartum depression: mapping the development from theory to practice. JMIR Res Protoc 2015 Oct 12;4(4):e120 [FREE Full text] [doi: 10.2196/resprot.4858] [Medline: 26476481]
34. Fonseca A, Monteiro F, Alves S, Gorayeb R, Canavarro MC. Be a mom, a web-based intervention to prevent postpartum depression: the enhancement of self-regulatory skills and its association with postpartum depressive symptoms. Front Psychol 2019 Feb 18;10:265 [FREE Full text] [doi: 10.3389/fpsyg.2019.00265] [Medline: 30873060]
35. Sheeber LB, Seeley JR, Feil EG, Davis B, Sorensen E, Kosty DB, et al. Development and pilot evaluation of an internet-facilitated cognitive-behavioral intervention for maternal depression. J Consult Clin Psychol 2012 Oct;80(5):739-749 [FREE Full text] [doi: 10.1037/a0028820] [Medline: 22663903]
36. Loughnan SA, Butler C, Sie AA, Grierson AB, Chen AZ, Hobbs MJ, et al. A randomised controlled trial of 'MUMentum postnatal': internet-delivered cognitive behavioural therapy for anxiety and depression in postpartum women. Behav Res Ther 2019 May;116:94-103. [doi: 10.1016/j.brat.2019.03.001] [Medline: 30877878]
37. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019 Aug;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]
38. Logsdon MC, Barone M, Lynch T, Robertson A, Myers J, Morrison D, et al. Testing of a prototype web based intervention for adolescent mothers on postpartum depression. Appl Nurs Res 2013 Aug;26(3):143-145. [doi: 10.1016/j.apnr.2013.01.005] [Medline: 23473677]

Abbreviations

AQoL-8D: Assessment of Quality of Life-8 Dimensions
CONSORT: Consolidated Standards of Reporting Trials
CSI-16: Couples Satisfaction Index-16
EPDS: Edinburgh Postnatal Depression Scale
MOS-SSS: Medical Outcomes Study Social Support Survey

Edited by R Kukafka, G Eysenbach; submitted 21.08.20; peer-reviewed by T Wentzel-Larsen, A Gulliver, T Donker; comments to author 21.10.20; revised version received 25.02.21; accepted 02.08.21; published 26.11.21.

Please cite as:
Kavanagh DJ, Connolly J, Fisher J, Halford WK, Hamilton K, Hides L, Milgrom J, Rowe H, Scuffham PA, White KM, Wittkowski A, Appleton S, Sanders D
The Baby Steps Web Program for the Well-Being of New Parents: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e23659
URL: https://www.jmir.org/2021/11/e23659
doi:10.2196/23659
PMID:34842534

Original Paper

Effectiveness of a Mobile Device–Based Resilience Training Program in Reducing Depressive Symptoms and Enhancing Resilience and Quality of Life in Parents of Children With Cancer: Randomized Controlled Trial

Yuanhui Luo[1], PhD; Wei Xia[2], PhD; Ankie Tan Cheung[3], PhD; Laurie Long Kwan Ho[3], PhD; Jingping Zhang[4], PhD; Jianhui Xie[5], BMed; Pin Xiao[6], BMed; Ho Cheung William Li[3], PhD

[1]School of Nursing, The University of Hong Kong, Hong Kong, China
[2]School of Nursing, Sun Yat-Sen University, Guangzhou, China
[3]Nethersole School of Nursing, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China
[4]Xiangya School of Nursing, Central South University, Changsha, China
[5]Department of Nursing, Hunan Children's Hospital, Changsha, China
[6]Department of Hematology, Hunan Children's Hospital, Changsha, China

Corresponding Author:
Ho Cheung William Li, PhD
Nethersole School of Nursing
The Chinese University of Hong Kong
Shatin, New Territories
Room 831
Esther Lee Building
Hong Kong
China
Phone: 86 39430889
Email: williamli@cuhk.edu.hk

Abstract

Background: Caring for children with cancer can be a stressful experience for parents and may have negative effects on their physical and psychological well-being. Although evidence has shown that resilience is associated with positive psychological well-being, few interventions have been specifically designed to enhance the resilience of parents of children with cancer.

Objective: The aim of this study is to examine the effectiveness of a mobile device–based resilience training program in reducing depressive symptoms and enhancing resilience and quality of life (QoL) in parents of children with cancer.

Methods: Parents of children diagnosed with cancer were recruited from the pediatric oncology wards of 3 tertiary hospitals in China. The participants were randomly assigned to either the experimental group (52/103, 50.5%) to undergo an 8-week mobile device–based resilience training program or to the control group (51/103, 49.5%) to receive an 8-week program of placebo information. The study outcomes included resilience, depressive symptoms, and QoL, as measured by the Connor–Davidson Resilience Scale, the Self-Rating Depression Scale, and the Short Form of the 6-Dimension Health Survey, respectively. All data were collected at baseline and at 2 and 6 months of follow-up. The data analysis followed the intention-to-treat principle. A generalized estimating equation was used to examine the effects of the intervention.

Results: The participants were mostly female (72/103, 69.9%), and their mean age was 33.6 (SD 5.2) years. The participants in the experimental group showed significantly higher levels of resilience (mean 67.96, SD 15.8 vs mean 58.27, SD 19.0; $P<.001$) and lower levels of depressive symptoms (mean 40.17, SD 9.9 vs mean 46.04, SD 10.9; $P<.001$) than those in the control group at 6 months of follow-up. The intervention showed statistically significant effects in improving resilience ($\beta=6.082$; $P=.01$) and decreasing depressive symptoms ($\beta=-2.772$; $P=.04$) relative to the control group. The QoL score in the experimental group was higher than that in the control group at 6 months of follow-up (mean 0.79, SD 0.2 vs mean 0.76, SD 0.3; $P=.07$); however, no statistically significant intervention effect was detected ($\beta=.020$; $P=.38$).

Conclusions: The mobile device–based resilience training program effectively enhanced resilience and alleviated depressive symptoms in parents of children with cancer. It is highly recommended that health care professionals incorporate this resilience training program when providing psychological care to parents of children with cancer.

Trial Registration: Clinical.Trials.gov NCT04038242; http://clinicaltrials.gov/ct2/show/NCT04038242

(J Med Internet Res 2021;23(11):e27639) doi:10.2196/27639

KEYWORDS

depressive symptoms; pediatric cancer; parents; quality of life; resilience; mobile phone

Introduction

Background

Approximately 300,000 new cases of childhood cancer are diagnosed annually worldwide [1]. The incidence of childhood cancer in China is similar to that of the world and has increased at a rate of 2.8% per year [2]. The diagnosis of cancer in children and long-term treatment are not only stressful events for the child but also disrupt the parents' daily lives, especially within the first year after the diagnosis, which has been described as the most chaotic moment of the parents' lives [3]. Parents must reorganize their family roles and routines, learn cancer-related information and care skills, prepare for overwhelming medical expenses, and manage intensive treatment regimes, each of which can cause tremendous stress and affect the parents' physical and psychological well-being [4].

Of all the negative psychological consequences, depressive feelings are one of the most common concerns reported by parents of children with cancer [5,6]. A recent meta-analysis found that the prevalence of depressive symptoms among parents of children with cancer varied from 7% to 91%, with a pooled prevalence of 28% [5]. A cross-sectional study of Chinese parents of children with leukemia found that as many as 77% of mothers and 42% of fathers had depressive symptoms [7]. Although the negative effects of parents' depressive symptoms on the quality of life (QoL) and well-being of both the parents and their children with cancer are widely acknowledged [8,9], little attention has been devoted to the prevention and alleviation of depressive symptoms in Chinese parents of children with cancer.

Resilience is the process of adapting well in the face of stress or adversity; it involves personal virtues and strengths that can be accessed and cultivated to achieve growth under detrimental conditions [10]. Evidence has shown that resilience is associated with psychological well-being in both clinical and nonclinical populations [11,12]. A strong negative relationship between resilience and depressive symptoms has also been found in parents of children with cancer [13]. It has been suggested that resilience can play an important role in protecting individuals from stress-related disorders [14]. Therefore, interventions that enhance resilience in the parents of children with cancer may influence their depressive symptoms and well-being.

A recent meta-analysis revealed that resilience training programs that aim to equip individuals with the resources and skills to navigate adversity and thrive in challenging environments had a moderately positive effect on the subjects' resilience [15]. These resilience training programs tend to include a combination of evidence-based training in areas such as cognitive strategies and mindfulness [15]. A review of the literature shows that most resilience training programs target students and patients with chronic illnesses [16]. No study has yet examined the effectiveness of resilience training programs in promoting the psychological well-being of parents of children with cancer. A study in the United States demonstrated the moderate effectiveness of a resilience training program delivered face-to-face to the parents of children with serious illnesses [17]. However, face-to-face training is time-consuming and requires frequent visits, which often leads to low compliance [15]. It has also been reported that some parents of children with cancer have less motivation to attend face-to-face training because of their busy schedules [18]. Mobile health (mHealth), which refers to health practices conducted via mobile devices, is increasingly common and has gained support from the World Health Organization for health promotion and treatment compliance [19,20]. It has also been suggested that internet- and mobile-based psychological interventions are cost-effective for both psychiatric and medical conditions [21]. The use of mobile apps to deliver resilience training has certain advantages; for example, health care professionals can give the participants remote support and feedback to promote their psychological well-being. Most importantly, training via mobile apps is more flexible and efficient than face-to-face interventions [22], particularly during a pandemic, when the delivery of face-to-face health care interventions may not be feasible. It is worth noting that with the rapid development and high use rate of WeChat (Tencent Holdings Limited) [23], such mobile apps may have the potential to deliver health interventions, increase adherence to training, and serve a large number of participants. This study aims to explore the effectiveness of a mobile device–based resilience training program in reducing depressive symptoms and enhancing the resilience and QoL of parents of children with cancer.

Theoretical Framework

The training program used in this study was guided by the resilience framework developed by Kumpfer [24]. The resilience framework generated 5 internal resilience factors: cognitive, emotional, spiritual, behavioral, and physical. Specifically, cognitive competency can help individuals think rationally and protect them from negative thoughts; emotional stability can help individuals recognize and deal reasonably with negative emotions; an individual's spiritual belief in his or her ability to improve the situation and achieve goals can help them survive difficult times; behavioral skills such as problem-solving can increase an individual's self-efficacy to tackle life's problems; and individuals with few physical problems may internalize

their physical strength and interpret themselves as being psychologically healthy [24]. These internal resilience factors represent fundamental elements that have been shown to be essential to cope effectively with life stressors and to adapt well in the face of adversity [10]. Therefore, by cultivating the skills and resources to foster these internal resilience factors, we expect that the intervention would promote parental resilience and improve parents' well-being.

Methods

Trial Design

This study was a 2-arm, parallel-group, randomized controlled trial (RCT) that adheres to the CONSORT (Consolidated Standards of Reporting Trials) guidelines [25]. This study was approved by the institutional review board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 19-436) and registered at ClinicalTrials.gov (NCT04038242).

Participants

From August 2019 to July 2020, the parents of children who received a diagnosis of cancer in the pediatric oncology wards of 3 tertiary hospitals in China and who met the following inclusion criteria were invited to participate in this study: (1) a child (aged 0-19 years) in whom cancer was diagnosed within the past year, (2) the ability to read Chinese and speak Mandarin, and (3) a smartphone with the WeChat app. Parents with cognitive impairments or physical disabilities identified from medical records and those who were participating in other psychological interventions or consultations were excluded.

Power analysis using G*Power 3.1 was performed to estimate the sample size [26]. According to a previous study that explored the effectiveness of a resilience training program in parents of children with serious illnesses [17], the effect size for resilience was 0.59. Thus, to predict the difference between the 2 groups at a 5% level of significance and a power of 0.8, 37 subjects were required for each group. Given a potential attrition rate of 20%, an additional 10 subjects were needed for each group. Therefore, the total sample size for this study was 94, with 47 subjects in each group.

Randomization and Blinding

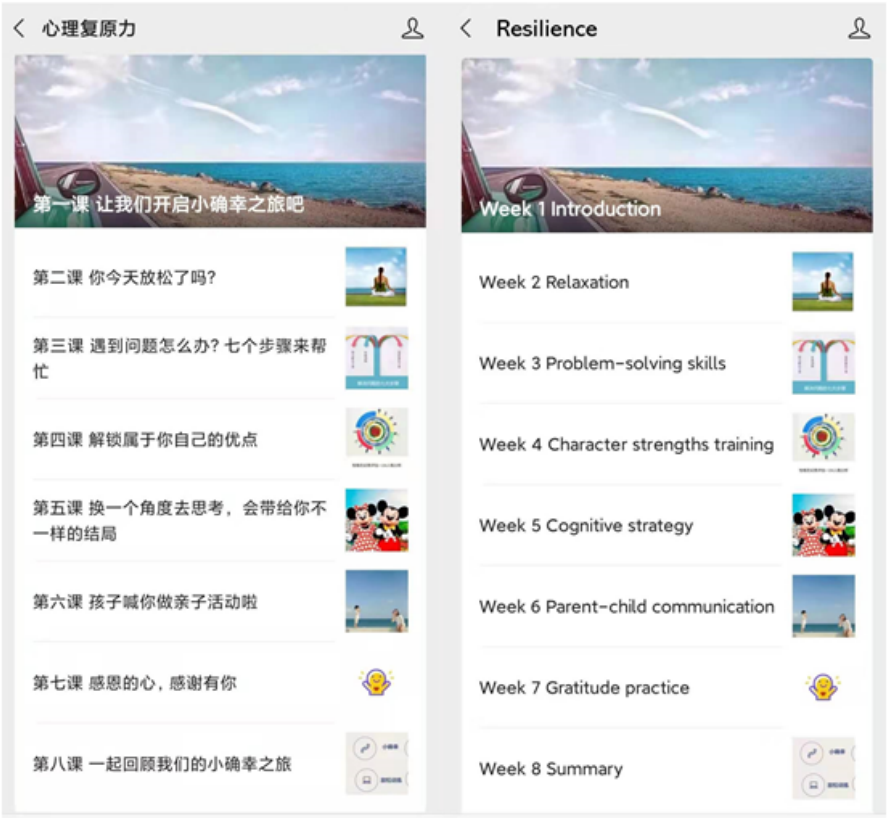
Simple randomization was used in this study. The participants were randomly allocated to the experimental group or control group in a 1:1 ratio. Random numbers were generated using IBM SPSS Statistics (version 25.0, IBM Corp) before recruitment. To guarantee allocation concealment, an independent researcher who was not involved in the recruitment matched the random numbers with the participants in chronological order of recruitment based on a WeChat official account follow list. Single blinding was used, and the research assistants who collected the data were blinded to the participants' group allocation.

Intervention

All participants were required to scan a quick response code to follow the WeChat official account at enrollment to participate in the intervention. The WeChat official account is the equivalent of a Facebook page and synthesizes functions, including tweet design and instant interaction communication [27]. The public can receive tweets after following the official WeChat account.

The participants in the experimental group participated in the resilience training program with 8 tweets that focused on the cultivation of internal resilience factors that enhance parental resilience (Figure 1). Each tweet session was constructed in 3 parts, including detailed skill training methods with pictures or short videos, examples of applying the skills when caring for children with cancer, and an assignment to reinforce the training (Multimedia Appendix 1). The participants were required to read the tweet sessions and finish the web-based assignments, which took approximately 15 minutes for each training session. Feedback on the assignment from a psychological consultant was sent back to the participants via WeChat within 1 working day. We sent the tweets weekly on each Saturday at 8 PM, and a reminder of the assignment was sent at the same time the next Wednesday to promote the participants' enrollment in training. To ensure that the intervention had an adequate effect on the outcome measures of this study, the dosage was assessed in terms of amount, frequency, and duration by a research committee that comprised a professor of psychology and an associate professor of psycho-oncology with extensive experience in psychological interventions, a professor of pediatric oncology and 2 senior pediatric oncology nurses with considerable clinical experience, and a research fellow with rich experience conducting resilience training programs. The feasibility and acceptability of the interventional dosage and content were evaluated in a 1-group pilot trial with 10 participants. Most of the participants commented that the length and frequency of the tweets were acceptable, and they were able to comprehend the content of the tweets. Hence, no changes were made to the content of the intervention. To control for potential contamination in the study, we set up the tweets to be sent only to specified participants, and the tweets could not be shared. Table 1 presents the detailed content of the mobile device–based resilience training program.

Figure 1. The total 8 tweets in the mobile device–based resilience training program.



The participants in the control group received 8 tweets that contained information from a caring manual distributed by the hospital to every child with cancer. The content of the tweets included oral care, symptom management, peripherally inserted central catheter maintenance, dietary guidance, medication care, knowledge about bone marrow biopsy, and infection prevention. The number and frequency of tweets were the same as those in the experimental group. No assignments or feedback were delivered to the participants in the control group.

Several strategies were adopted to ensure the integrity of the intervention. First, a research protocol was developed, and the interventions strictly followed the protocol. Second, the number of tweets and feedback sent to each participant was recorded in Microsoft Excel (Microsoft Corporation) to monitor the delivery of the interventions. In addition, team meetings were held monthly to evaluate the quality of the intervention's implementation.

Table 1. Content of the mobile device–based resilience training program.

Sessions or tweets and internal resilience factors	Details		
	Objectives	Core content	Assignment
Week 1: introduction			
Emotional	• To help the participants understand the purpose of the intervention • To help the participants initially understand the intervention content	Definition of resilience, purpose of the study, science of the intervention, general content of all tweets, and encouragement to cultivate positive emotions in daily life	The 3 good things practice: write down 3 things that go well for you that day and reflect on why they went well
Week 2: relaxation			
Emotional and physical	• To practice relaxation techniques • To reduce the participants' stress	Definition of meditation, science and specific practice methods of meditation exercise, and guided meditation audio	Guided breath awareness: sit quietly and be aware of your breath and exercise following the guided meditation audio
Week 3: problem-solving skills			
Behavioral	• To learn problem-solving skills • To cultivate a positive coping style	Introduction and science of the problem-solving therapy, specific 7 steps to solve problems, and examples of solving problems encountered in caring for children with cancer	Problem-solving practice: complete the problem-solving worksheet according to the instructions
Week 4: character strength training			
Cognitive	• To help the participants understand their own character strengths • To increase confidence in dealing with difficulties in life	Definition of character strength, establishment of character strength assessment system, and science and specific methods of character strength training	Character strength training: finish web-based character strength assessment to find your top 5 strengths. Use your strengths in caring for your child
Week 5: cognitive strategies			
Cognitive and emotional	• To help the participants reframe negative or unhelpful cognition • To learn emotion management skills	Recognize and accept emotions, definition of cognition, science of cognition restructuring, examples of cognition restructuring scenarios from parents of children with cancer, and emotion management strategies	Cognition restructuring practice: complete the cognitive reconstructing worksheet according to the instructions
Week 6: parent–child communication			
Behavioral	• To promote effective communication between parent and child • To help the participants build a good relationship with their child	Ways to achieve effective parent–child communication, examples of good communication from parents of children with cancer, and tips for managing children's emotions	Parent–child activity: accompany the child for an appropriate activity and apply the communication skills in the activity
Week 7: gratitude practice			
Spiritual	• To help the participants cultivate positive beliefs • To help the participants attain personal growth	Manifestations, benefits, science and specific exercise of gratitude; steps to make a realizable goal; and examples of goal setting from parents of children with cancer	Gratitude activity: write a gratitude letter or keep a gratitude diary or prepare a gratitude card
Week 8: summary			
Cognitive	• To review the learned skills and related assignments • To help the participants make a plan to reinforce the learned skills in their future life	Simple summary for each tweet, emphasis of the learned resilience skills and related assignments, and encourage the participants to make a plan to practice the learned skills	Plan-making: make a plan to practice all skills and choose a favorite exercise to keep up

Outcomes and Measures

At baseline, the participants' demographic information and their children's clinical characteristics were collected using a demographic information sheet. The questionnaire requested information about the participants' gender, age, marital status, educational attainment, and monthly income. The participants also reported their child's age, type of cancer, time since diagnosis, and risk stratification of cancer. In resilience intervention studies, it has been suggested that a posttest at the end of the training program and a follow-up assessment at 6 months are necessary to evaluate the efficacy of the interventions [28]. Hence, the outcomes in this study were assessed at baseline and at 2 and 6 months after the intervention began.

The primary outcome was resilience, which was measured at 6 months using the Connor–Davidson Resilience Scale, a self-rated scale designed by Connor and Davidson [29]. The Chinese version was translated and tested by Yu and Zhang [30] and has shown good reliability and satisfactory validity. It consists of 25 items and is scored on a 5-point Likert scale (0-4). The total score ranges from 0 to 100, with a higher score reflecting a higher level of resilience. It has been widely used in clinical practice and treatment-outcome research [31].

The secondary outcomes were depressive symptoms and QoL at 6 months of follow-up and resilience, depressive symptoms, and QoL at the 2-month follow-up visit. The Self-Rating Depression Scale was used to measure participants' depressive symptoms during the previous week [32]. It includes 20 items, and half of the items are scored in reverse. The frequency of the symptom in each item was evaluated on a 4-point Likert scale (1=never or seldom, 2=sometimes, 3=often, and 4=always). The total score ranges from 20 to 80, and a higher total score indicates worse depressive symptoms. The Chinese version of this scale has shown good reliability and satisfactory internal consistency [33].

The Short Form of the 6-Dimension Health Survey was used to measure the participants' QoL. It was developed by Brazier et al [34] and is derived from the 36-item Short Form Health Survey. Hong Kong scholars translated it into Chinese and produced a scoring algorithm [35]. This scale is a 6-dimension health survey that assesses physical functioning, role limitations, social functioning, pain, mental health, and vitality. The total score based on the scoring algorithm ranges from 0.32 for worst health to 1 for full health. The scale has been shown to be reliable and valid in a Chinese population [36].

Data Collection

To identify potential participants, posters with information about the study were posted on the notice board in the hospitals' pediatric oncology wards and outpatient clinics. Parents who were willing to participate scanned a quick response code on the poster to complete a simple application form. The research assistant then contacted the person who filled out the application form after assessing their eligibility. The research assistant also visited each ward and invited parents of children with cancer to participate. The study details, including purpose, procedures, and potential benefits and harm, were explained, and the parents were given the option of participation or refusal. The parents were also told that their participation was voluntary and that they had the right to withdraw at any time without reprisal. Written informed consent was then obtained from parents who agreed to participate. The baseline and 2-month follow-up data were collected using written questionnaires in the wards of the hospitals. Outcome measures at the 6-month follow-up were collected using web-based questionnaires.

Statistical Analysis

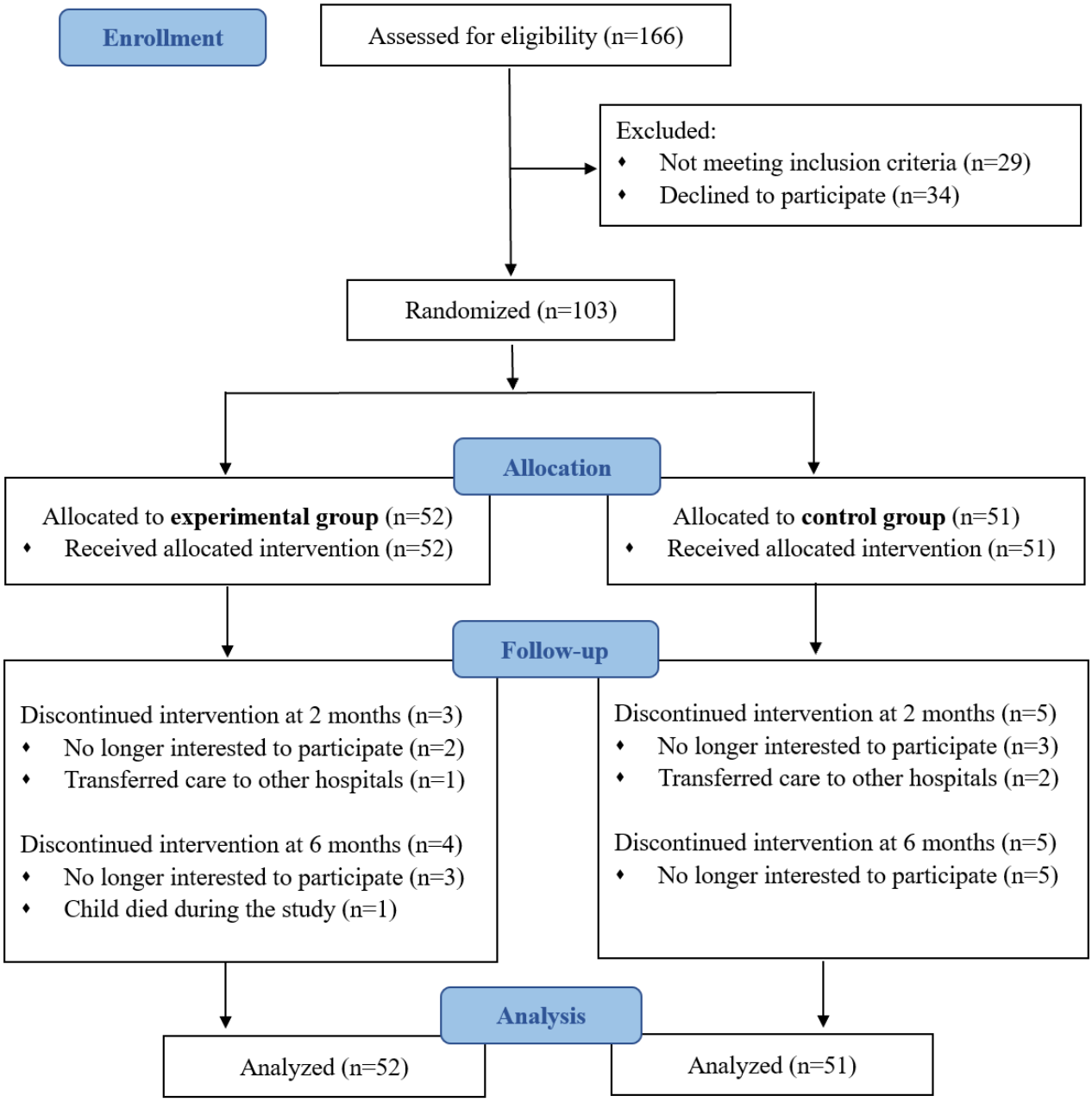
IBM SPSS Statistics 25.0 (IBM Corp) was used for data analysis. The intention-to-treat principle was applied, and the participants were analyzed according to their initial group assignments. Missing data were handled with multiple imputation using the Markov chain Monte Carlo method. Complete case analysis was conducted to test the robustness of multiple imputation. Descriptive statistics were applied to calculate the mean and SD for continuous data and the frequency and percentage for categorical data. The range and median values for the study outcomes were displayed as a complement. The differences in the baseline characteristics and outcomes between the experimental and control groups were examined using a chi-square test or Fisher exact test for categorical variables and 2-tailed independent-sample *t* test and Wilcoxon signed-rank test for continuous variables. A generalized estimating equation was used to estimate the effects of the intervention on all outcomes with repeated measures [37]. Statistical significance was set at $P<.05$.

Results

Participants Characteristics

A total of 103 participants were recruited; of the 103 participants, the experimental and control groups had 52 (50.5%) and 51 (49.5%) participants, respectively. Figure 2 presents the flowchart of the study. The overall attrition rate was 16.5%, and no significant differences were found in the baseline characteristics between the participants who withdrew and those who remained.

Figure 2. A CONSORT (Consolidated Standards of Reporting Trial) flow diagram.



The participants were mostly female (72/103, 69.9%) and married (100/103, 97.1%), and their mean age was 33.6 (SD 5.2) years. Approximately one-third (29/103, 28.2%) had attained tertiary education, and half (52/103, 50.5%) had a monthly income of <Chinese ¥3000 (US $427.35). Approximately two-thirds (69/103, 67%) of the participants' children received a diagnosis of a hematologic tumor, and the remaining one-third (34/103, 33%) had a solid tumor. The children's mean age was 5.9 (SD 3.9) years, and most had received their diagnosis <6 months earlier (85/103, 82.5%). No significant differences were detected in the baseline characteristics between the experimental and control groups (P>.05); details are given in Table 2.

Table 2. Baseline characteristics of the participants in the experimental and control groups (N=103).

Characteristics	Experimental (n=52)		Control (n=51)		Chi-square (*df*)	*t* test (*df*)[a]	*P* value
	Participant, n (%)	Values, mean (SD)	Participant, n (%)	Values, mean (SD)			
Gender		N/A[b]		N/A		N/A	
Female	35 (67)		37 (73)		0.3 (1)		.56
Male	17 (33)		14 (23)		0.3 (1)		.56
Marital status		N/A		N/A		N/A	
Married	50 (96)		50 (98)		0.3 (1)		.99[c]
Divorced	2 (4)		1 (2)		0.3 (1)		.99[c]
Educational attainment		N/A		N/A		N/A	
Primary school	5 (10)		3 (6)		1.1 (2)		.60[c]
High school	31 (60)		35 (69)		1.1 (2)		.60[c]
College	16 (31)		13 (25)		1.1 (2)		.60[c]
Monthly income (Chinese ¥; US$)		N/A		N/A		N/A	
<3000 (<427.35)	24 (46)		28 (55)		1.2 (2)		.54
3000-5000 (427.35-712.25)	17 (33)		16 (31)		1.2 (2)		.54
>5000 (>712.25)	11 (21)		7 (14)		1.2 (2)		.54
Type of cancer		N/A		N/A		N/A	
Hematology tumor	38 (73)		31 (61)		1.8 (1)		.19
Solid tumor	14 (27)		20 (39)		1.8 (1)		.19
Time since diagnosis (months)		N/A		N/A		N/A	
0-6	44 (85)		41 (80)		0.3 (1)		.57
7-12	8 (15)		10 (20)		0.3 (1)		.57
Risk stratification of cancer		N/A		N/A		N/A	
Low	18 (35)		17 (33)		1.0 (2)		.61
Intermediate	22 (42)		18 (35)		1.0 (2)		.61
High	12 (23)		16 (31)		1.0 (2)		.61
Parents' age in years	N/A	33.92 (5.4)	N/A	33.22 (5.0)	N/A	−0.691 (101)	.49
Children's age in years	N/A	5.48 (3.7)	N/A	6.41 (3.6)	N/A	1.278 (101)	.20

[a]The significance of *t* test is 2 tailed.

[b]N/A: not applicable.

[c]Fisher exact test.

Comparisons Between Experimental and Control Groups

Table 3 presents comparisons of resilience, depressive symptoms, and QoL between the 2 groups at each measuring point. Compared with the participants in the control group, those in the experimental group showed statistically higher levels of resilience at 2 months (mean 61.90, SD 14.6 vs mean 69.35, SD 13.4; *P*=.005) and 6 months (mean 58.27, SD 19.0 vs mean 67.96, SD 15.8; *P*<.001) and lower levels of depressive symptoms at 2 months (mean 44.66, SD 8.0 vs mean 40.40, SD 9.1; *P*=.009) and 6 months (mean 46.04, SD 10.9 vs mean 40.17, SD 9.9; *P*<.001). Although the QoL scores in the control group were lower than those in the experimental group, no significant differences were found between the 2 groups at 2 months (mean 0.75, SD 0.1 vs mean 0.77, SD 0.2; *P*=.11) and 6 months (mean 0.76, SD 0.3 vs mean 0.79, SD 0.2; *P*=.07).

Table 3. Comparison of resilience, depressive symptoms, and quality of life in the experimental and control groups (N=103).

Outcomes measures	Experimental (n=52), mean (SD)	Control (n=51), mean (SD)	*t* test (*df*)[a]	*P* value	Experimental (n=52), median (range)	Control (n=51), median (range)	*Z*	*P* value
Resilience								
Baseline	63.48 (15.1)	61.73 (14.7)	–0.598 (101)	.55	63 (32-96)	62 (18-84)	–0.376	.71
2 months	69.35 (13.4)	61.90 (14.6)	–2.673 (101)	.005	69 (31-96)	62 (12-94)	–2.846	.004
6 months	67.96 (15.8)	58.27 (19.0)	–3.521 (101)	<.001	66 (41-100)	58 (22-87)	–3.102	.002
Depressive symptoms								
Baseline	45.40 (7.7)	44.16 (7.1)	–0.852 (101)	.40	45 (30-67)	44 (29-70)	–0.740	.46
2 months	40.40 (9.1)	44.66 (8.0)	2.554 (101)	.009	39 (24-62)	45 (24-71)	–2.665	.008
6 months	40.17 (9.9)	46.04 (10.9)	3.467 (101)	<.001	40 (25-63)	45 (31-64)	–3.025	.002
Quality of life								
Baseline	0.77 (0.1)	0.76 (0.1)	–0.191 (101)	.85	0.82 (0.41-1.00)	0.78 (0.44-0.96)	–0.801	.42
2 months	0.77 (0.2)	0.75 (0.1)	–1.589 (101)	.11	0.82 (0.49-0.96)	0.76 (0.45-1.00)	–1.456	.15
6 months	0.79 (0.2)	0.76 (0.3)	–1.791 (101)	.07	0.81 (0.47-1.00)	0.79 (0.48-1.00)	–1.340	.18

[a]The significance of *t* test is 2-tailed.

Generalized Estimating Equation Model

There were 16.5% (17/103) of cases with missing data on the variables during follow-up, including resilience, depressive symptoms, and QoL. Table 4 shows the results of the generalized estimating equation model for study outcomes under the intention-to-treat and complete case analyses. Similar results between the 2 analyses reflected the robustness of multiple imputation. Statistically significant main effects were found for the intervention in promoting resilience (β=6.082; *P*=.01) and decreasing depressive symptoms (β=−2.772; *P*=.04), whereas no significant intervention effect was observed for QoL (β=.020; *P*=.38). In addition, significant group-by-time interaction effects were detected for resilience at 2 months (β=5.812; *P*=.01) and 6 months (β=7.167; *P*=.01) and for depressive symptoms at 2 months (β=−5.553; *P*<.001) and 6 months (β=−6.504; *P*<.001). No significant interaction effects were observed for QoL at 2 months (β=.021; *P*=.42) or 6 months (β=.023; *P*=.43).

Table 4. A generalized estimating equation model for resilience, depressive symptoms, and quality of life.

Outcome measures	Intention-to-treat (N=103)		Complete case (N=86)	
	β (SE; 95% CI)	*P* value	β (SE; 95% CI)	*P* value
Resilience				
Main effect[a]	6.082 (2.360; 1.455 to 10.709)	.01	6.055 (2.652; 0.858 to 11.253)	.02
Group×time 1[b,c]	5.812 (2.363; 1.177 to 10.448)	.01	6.373 (2.226; 2.011 to 10.736)	.005
Group×time 2[b,c]	7.167 (2.921; 1.436 to 12.899)	.01	7.605 (2.739; 2.236 to 12.974)	.004
Depressive symptoms				
Main effect	−2.772 (1.354; −5.427 to −0.117)	.04	−2.454 (1.581; −5.553 to 0.644)	.12
Group×time 1	−5.553 (1.233; −7.971 to −3.135)	<.001	−5.403 (1.177; −7.710 to −3.096)	<.001
Group×time 2	−6.504 (1.592; −9.633 to −3.375)	<.001	−7.251 (1.417; −10.028 to −4.474)	<.001
Quality of life				
Main effect	.020 (0.023; −0.025 to 0.064)	.38	.014 (0.025; −0.035 to 0.063)	.58
Group×time 1	.021 (0.026; −0.030 to 0.073)	.42	.021 (0.027; −0.032 to 0.074)	.45
Group×time 2	.023 (0.030; −0.035 to 0.082)	.43	.021 (0.027; −0.032 to 0.073)	.44

[a]Referred to control group.

[b]Referred to control group and baseline.

[c]Time 1 and time 2 refer to 2 and 6 months of follow-up, respectively.

Discussion

Principal Findings

This RCT addressed an important health issue related to the consequences of children's cancer on their parents' psychological well-being and QoL, an area that has not been well explored. Most importantly, this study is original and helps to clarify the effectiveness of a mobile device–based resilience training program in reducing depressive symptoms and enhancing resilience and QoL in Chinese parents of children with cancer. To our knowledge, this study is the first RCT that combines mHealth and resilience training to improve mental health in the parents of children with cancer.

This study demonstrated that over a 2-month period, this mobile device–based resilience training program was sufficient to enhance resilience in the parents of children with cancer. The findings were in accordance with those of a previous study conducted in students using a computer device compared with educational attention control [38]. Guided by the resilience framework of Kumpfer [24], our program provided training in skills to foster internal resilience factors and enhance the resilience levels of parents of children with cancer. For example, the practice of recording 3 good things every day could encourage the participants to focus on positive things so as to cultivate positive emotions and optimism related to internal emotional characteristics [39]. Cognitive reframing training, which usually involves internal cognitive characteristics, could help the participants view the situation from various perspectives and change their negative thoughts into positive ones, thus allowing them to deal with trouble in cognition [40]. Moreover, to foster internal spiritual characteristics, the participants were encouraged to express their gratitude in various ways to those who supported them. Such gratitude practices have been shown to cultivate participants' positive beliefs and help them attain personal growth [41]. In addition, the difference in resilience between the experimental and control groups was still significant after 6 months of follow-up, which indicates that the effectiveness of a mobile device–based resilience training program can be sustained over an extended period. Increased proficiency in skills use may play a role in maintaining resilience levels and help parents better navigate adversity [42].

The mobile device–based resilience training program significantly alleviated depressive symptoms in parents of children with cancer. The depressive symptoms score increased continuously across the study for the parents in the control group, whereas the score for the parents in the experimental group showed an opposite trend. These findings were consistent with those of previous resilience training programs for patients with chronic illnesses [16] and contrary to the changes in resilience in this study. The negative relationship between resilience and depressive symptoms has been widely acknowledged [33,43], and our findings further support this relationship in parents of children with cancer. It has been noted that resilience can protect individuals from stress-related disorders such as depressive symptoms via positive cognition and enhanced self-efficacy [44]. Individuals with high levels of resilience are more likely to believe that they have the skills to deal with an adverse situation and to appraise the situation as a challenge to overcome [45]. In this study, problem-solving skills were taught to help parents deal with the issues they encountered in caring for their children with cancer. Combined with character strength training, parents were expected to use their cultivated strengths to solve related problems in daily life. Hence, the learned skills that could increase parents' confidence in coping with their children's cancer may have a positive effect on alleviating their depressive symptoms. Furthermore, evidence has shown that relaxation techniques and strong relationships

are beneficial for ameliorating depressive symptoms [46,47]. The guided breath awareness practice and parent–child communication training included in the resilience training program may also play a role in reducing depressive symptoms in parents of children with cancer.

Although the QoL scores in the experimental group were higher than those in the control group, no significant effects were detected. The improvements in resilience and alleviation of depressive symptoms did not significantly promote QoL in the parents of children with cancer, possibly because of the relatively short study period. It has been well demonstrated that a change in QoL usually takes a longer time than behavior changes or mental health improvement [48]. Previous studies generally included 3 months of follow-up to explore the effect of a resilience training program on QoL, and the pooled results of the systematic review showed no significant improvement [16]. It is recommended that future studies add further evidence on the long-term effects of a mobile device–based resilience training program on QoL in parents of children with cancer. In addition, QoL is regarded as a multidimensional concept that is influenced by various factors other than resilience and depressive symptoms [49,50]. The skills trained in this study focused mainly on promoting the parents' psychological well-being, and strategies that target the cultivation of social resources and improvement of physical health are worth considering in future interventions. Evidence has shown that COVID-19 is a stressful event for the general public and can have an impact on their QoL [51]. Given that the study period covered the first wave of COVID-19 in China, the pandemic might have had an impact on the parental QoL and weakened the intervention effects. However, the efficacy in enhancing resilience and reducing depressive symptoms indicated that the learned skills in resilience training could help parents adapt to adversities, including the pandemic.

Limitations

There were 2 limitations to this study. First, we only examined the effects of the mobile device–based resilience training program on parents of children with cancer for 6 months. Whether the improvement in resilience and the reduction in depressive symptoms could be sustained or whether the changes in QoL could show a significant difference over the long term should be further explored. Second, because of the small sample size, we did not perform a statistical assessment to identify the potential impacts of social demographics and children's clinical characteristics (ie, time since diagnosis) on parents' outcomes. Nevertheless, we conducted a cross-sectional study to explore the relationships between resilience and QoL in Chinese parents of children with cancer [52] before this RCT. The results showed that the time since diagnosis did not correlate with parents' resilience and QoL.

Implications for Future Practice and Research

Given the increasing incidence of childhood cancer and considerable stress in caring for these children [1], it is critical in clinical practice to provide effective psychological interventions for parents of children with cancer. Our findings indicate that our mobile device–based resilience training program was efficient in enhancing resilience and reducing depressive symptoms. As a potential preventive strategy for stress-related disorders [14], the application of a resilience training program is recommended at an early stage when parents first face the diagnosis of their child's cancer. The learned skills would then promote parental resilience and help them adapt more quickly to their child's cancer and deal with the adversity without the harassment of stress-related disorders. Considering the effectiveness and convenience of the mobile device–based resilience training program, health care professionals could use popular social media apps such as WeChat, Facebook, and WhatsApp to implement such programs to benefit more parents of children with cancer via mHealth care. As objective biometric data can complement the subjective evidence of the effects of the mobile device–based resilience training program, relative biometric outcomes should be assessed in future studies. In addition, given that changes in children's health conditions might affect the parents' outcomes, it is recommended that future studies also assess the health condition of children with cancer over time. Finally, the parental physical symptoms, such as insomnia, could be good indicators of the effectiveness of resilience training programs, which should be assessed in future studies.

Conclusions

Our mobile device–based resilience training program was developed under the resilience framework of Kumpfer [24] and included 8 tweets and assignments to train skills that foster internal resilience factors. The parents of children with cancer who participated in this program revealed higher levels of resilience and fewer depressive symptoms than the control subjects. It is highly recommended that health care professionals incorporate this resilience training program when providing psychological care to parents of children with cancer.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English text version of 8 tweets in the mobile device–based resilience training program.
[DOCX File , 32 KB - jmir_v23i11e27639_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 726 KB - jmir_v23i11e27639_app2.pdf]

References

1. Steliarova-Foucher E, Colombet M, Ries LA, Moreno F, Dolya A, Bray F, IICC-3 Contributors. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol 2017 Jun;18(6):719-731 [FREE Full text] [doi: 10.1016/S1470-2045(17)30186-9] [Medline: 28410997]
2. Zheng R, Peng X, Zeng H, Zhang S, Chen T, Wang H, et al. Incidence, mortality and survival of childhood cancer in China during 2000-2010 period: a population-based study. Cancer Lett 2015 Jul 28;363(2):176-180. [doi: 10.1016/j.canlet.2015.04.021] [Medline: 25917566]
3. Mu PF, Lee MY, Sheng CC, Tung PC, Huang LY, Chen YW. The experiences of family members in the year following the diagnosis of a child or adolescent with cancer: a qualitative systematic review. JBI Database System Rev Implement Rep 2015 Jun 12;13(5):293-329. [doi: 10.11124/jbisrir-2015-1698] [Medline: 26455612]
4. Witt WP, Litzelman K, Wisk LE, Spear HA, Catrine K, Levin N, et al. Stress-mediated quality of life outcomes in parents of childhood cancer and brain tumor survivors: a case-control study. Qual Life Res 2010 Sep 15;19(7):995-1005. [doi: 10.1007/s11136-010-9666-9] [Medline: 20473638]
5. van Warmerdam J, Zabih V, Kurdyak P, Sutradhar R, Nathan PC, Gupta S. Prevalence of anxiety, depression, and posttraumatic stress disorder in parents of children with cancer: a meta-analysis. Pediatr Blood Cancer 2019 Jun 28;66(6):e27677. [doi: 10.1002/pbc.27677] [Medline: 30816008]
6. Collins ML, Bregman J, Ford JS, Shields CL. Depression, anxiety, and stress in parents of patients with retinoblastoma. Am J Ophthalmol 2019 Nov;207:130-143. [doi: 10.1016/j.ajo.2019.05.020] [Medline: 31163135]
7. Chen J, Liu Y, Cai Q, Liu Y, Wang T, Wang J, et al. Depression in parents of children with leukemia in southern China accompanied by the prevalence of type D personality. Support Care Cancer 2014 May 11;22(5):1277-1286. [doi: 10.1007/s00520-013-2082-2] [Medline: 24326951]
8. Bakula DM, Sharkey CM, Perez MN, Espeleta HC, Gamwell KL, Baudino M, et al. Featured article: the relationship between parent and child distress in pediatric cancer: a meta-analysis. J Pediatr Psychol 2019 Nov 01;44(10):1121-1136. [doi: 10.1093/jpepsy/jsz051] [Medline: 31260071]
9. Ilic V, Nikitovic M, Maric G, Jovanovic A, Paripovic L, Bokun J, et al. Assessment of health-related quality of life among parents of children with solid tumors in Serbia. Support Care Cancer 2020 Nov 10;28(11):5109-5115. [doi: 10.1007/s00520-020-05348-6] [Medline: 32040636]
10. Richardson GE. The metatheory of resilience and resiliency. J Clin Psychol 2002 Mar;58(3):307-321. [doi: 10.1002/jclp.10020] [Medline: 11836712]
11. Kirby JS, Butt M, Esmann S, Jemec GB. Association of resilience with depression and health-related quality of life for patients with hidradenitis suppurativa. JAMA Dermatol 2017 Dec 01;153(12):1263-1269 [FREE Full text] [doi: 10.1001/jamadermatol.2017.3596] [Medline: 29117300]
12. Li ZS, Hasson F. Resilience, stress, and psychological well-being in nursing students: a systematic review. Nurse Educ Today 2020 Apr 19;90:104440. [doi: 10.1016/j.nedt.2020.104440] [Medline: 32353643]
13. Ye ZJ, Qiu HZ, Li PF, Liang MZ, Wang SN, Quan XM. Resilience model for parents of children with cancer in mainland China-an exploratory study. Eur J Oncol Nurs 2017 Apr;27:9-16. [doi: 10.1016/j.ejon.2017.01.002] [Medline: 28279395]
14. Kalisch R, Baker DG, Basten U, Boks MP, Bonanno GA, Brummelman E, et al. The resilience framework as a strategy to combat stress-related disorders. Nat Hum Behav 2017 Nov 16;1(11):784-790. [doi: 10.1038/s41562-017-0200-8] [Medline: 31024125]
15. Joyce S, Shand F, Tighe J, Laurent SJ, Bryant RA, Harvey SB. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. BMJ Open 2018 Jun 14;8(6):e017858 [FREE Full text] [doi: 10.1136/bmjopen-2017-017858] [Medline: 29903782]
16. Leppin AL, Bora PR, Tilburt JC, Gionfriddo MR, Zeballos-Palacios C, Dulohery MM, et al. The efficacy of resiliency training programs: a systematic review and meta-analysis of randomized trials. PLoS One 2014 Oct 27;9(10):e111420 [FREE Full text] [doi: 10.1371/journal.pone.0111420] [Medline: 25347713]
17. Yi-Frazier JP, Fladeboe K, Klein V, Eaton L, Wharton C, McCauley E, et al. Promoting Resilience in Stress Management for Parents (PRISM-P): an intervention for caregivers of youth with serious illness. Fam Syst Health 2017 Sep;35(3):341-351 [FREE Full text] [doi: 10.1037/fsh0000281] [Medline: 28541057]
18. Applebaum AJ, Breitbart W. Care for the cancer caregiver: a systematic review. Palliat Support Care 2012 Oct 10;11(3):231-252. [doi: 10.1017/s1478951512000594]
19. Krishna R. The impact of health information technology on the doctor-patient relationship in child and adolescent psychiatry. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):67-75. [doi: 10.1016/j.chc.2016.07.007] [Medline: 27837943]
20. WHO Global Observatory for eHealth. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
21. Zhang MW, Ho RC. Moodle: the cost effective solution for internet cognitive behavioral therapy (I-CBT) interventions. Technol Health Care 2017 Feb 21;25(1):163-165. [doi: 10.3233/thc-161261] [Medline: 27689560]

22. Uthman OA, Nduka CU, Abba M, Enriquez R, Nordenstedt H, Nalugoda F, et al. Comparison of mHealth and face-to-face interventions for smoking cessation among people living with HIV: meta-analysis. JMIR mHealth uHealth 2019 Jan 07;7(1):e203 [FREE Full text] [doi: 10.2196/mhealth.9329] [Medline: 30617044]
23. Number of monthly active WeChat users from 2nd quarter 2011 to 2nd quarter 2020. Statista. 2021. URL: https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/#statisticContainer [accessed 2021-11-05]
24. Kumpfer KL. Factors and processes contributing to resilience: the resilience framework. In: Resilience and Development: Positive Life Adaptations. Dordrecht: Springer; 1999.
25. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
26. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
27. WeChat official account: a simple guide. Walk The Chat. 2019. URL: https://walkthechat.com/wechat-official-account-simple-guide/#wechat-official-acct [accessed 2021-11-05]
28. Chmitorz A, Kunzler A, Helmreich I, Tüscher O, Kalisch R, Kubiak T, et al. Intervention studies to foster resilience - a systematic review and proposal for a resilience framework in future intervention studies. Clin Psychol Rev 2018 Feb;59:78-100 [FREE Full text] [doi: 10.1016/j.cpr.2017.11.002] [Medline: 29167029]
29. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003 Sep 02;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
30. Yu X, Zhang J. Factor analysis and psychometric evaluation of the Connor-Davidson Resilience Scale (CD-RISC) with Chinese people. Soc Behav Pers 2007 Jan 01;35(1):19-30. [doi: 10.2224/sbp.2007.35.1.19]
31. Kim GM, Lim JY, Kim EJ, Park SM. Resilience of patients with chronic diseases: a systematic review. Health Soc Care Community 2019 Jul 20;27(4):797-807. [doi: 10.1111/hsc.12620] [Medline: 30027595]
32. Zung WW. A self-rating depression scale. Arch Gen Psychiatry 1965 Jan 01;12:63-70. [doi: 10.1001/archpsyc.1965.01720310065008] [Medline: 14221692]
33. Ye ZJ, Guan HJ, Wu LH, Xiao MY, Luo DM, Quan XM. Resilience and psychosocial function among mainland chinese parents of children with cancer: a cross-sectional survey. Cancer Nurs 2015;38(6):466-474. [doi: 10.1097/NCC.0000000000000220] [Medline: 25629894]
34. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar;21(2):271-292. [doi: 10.1016/s0167-6296(01)00130-8] [Medline: 11939242]
35. McGhee SM, Brazier J, Lam CL, Wong LC, Chau J, Cheung A, et al. Quality-adjusted life years: population-specific measurement of the quality component. Hong Kong Med J 2011 Dec;17(6):17-21 [FREE Full text] [Medline: 22147354]
36. Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. Value Health 2008 Mar;11(2):295-303 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00233.x] [Medline: 18380642]
37. Ma Y, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. Reg Anesth Pain Med 2012;37(1):99-105 [FREE Full text] [doi: 10.1097/AAP.0b013e31823ebc74] [Medline: 22189576]
38. Rose RD, Buckey Jr JC, Zbozinek TD, Motivala SJ, Glenn DE, Cartreine JA, et al. A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. Behav Res Ther 2013 Feb;51(2):106-112. [doi: 10.1016/j.brat.2012.11.003] [Medline: 23262118]
39. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
40. Kuhlthau KA, Luberto CM, Traeger L, Millstein RA, Perez GK, Lindly OJ, et al. A virtual resiliency intervention for parents of children with autism: a randomized pilot trial. J Autism Dev Disord 2020 Jul 21;50(7):2513-2526 [FREE Full text] [doi: 10.1007/s10803-019-03976-4] [Medline: 30900195]
41. Martin F, Clyne W, Pearce G, Turner A. Self-management support intervention for parents of children with developmental disorders: the role of gratitude and hope. J Child Fam Stud 2019 Feb 13;28(4):980-992. [doi: 10.1007/s10826-018-01308-1]
42. Rosenberg AR, Baker KS, Syrjala KL, Back AL, Wolfe J. Promoting resilience among parents and caregivers of children with cancer. J Palliat Med 2013 Jun;16(6):645-652 [FREE Full text] [doi: 10.1089/jpm.2012.0494] [Medline: 23646887]
43. Rosenberg AR, Wolfe J, Bradford MC, Shaffer ML, Yi-Frazier JP, Curtis JR, et al. Resilience and psychosocial outcomes in parents of children with cancer. Pediatr Blood Cancer 2014 Mar 19;61(3):552-557 [FREE Full text] [doi: 10.1002/pbc.24854] [Medline: 24249426]
44. Horn SR, Feder A. Understanding resilience and preventing and treating PTSD. Harv Rev Psychiatry 2018;26(3):158-174. [doi: 10.1097/HRP.0000000000000194] [Medline: 29734229]
45. Southwick SM, Charney DS. The science of resilience: implications for the prevention and treatment of depression. Science 2012 Oct 05;338(6103):79-82. [doi: 10.1126/science.1222942] [Medline: 23042887]
46. Stonnington CM, Darby B, Santucci A, Mulligan P, Pathuis P, Cuc A, et al. A resilience intervention involving mindfulness training for transplant patients and their caregivers. Clin Transplant 2016 Nov 14;30(11):1466-1472. [doi: 10.1111/ctr.12841] [Medline: 27618687]

47. Chan PT, Doan SN, Tompson MC. Stress generation in a developmental context: the role of youth depressive symptoms, maternal depression, the parent-child relationship, and family stress. J Fam Psychol 2014 Feb;28(1):32-41. [doi: 10.1037/a0035277] [Medline: 24377729]
48. Li WH, Ho KY, Lam KK, Lam HS, Chui SY, Chan GC, et al. Adventure-based training to promote physical activity and reduce fatigue among childhood cancer survivors: a randomized controlled trial. Int J Nurs Stud 2018 Jul;83:65-74. [doi: 10.1016/j.ijnurstu.2018.04.007] [Medline: 29689482]
49. Rensen N, Steur LM, Schepers SA, Merks JH, Moll AC, Kaspers GJ, et al. Gender-specific differences in parental health-related quality of life in childhood cancer. Pediatr Blood Cancer 2019 Jul 27;66(7):e27728. [doi: 10.1002/pbc.27728] [Medline: 30916456]
50. Modanloo S, Rohani C, Farahani AS, Vasli P, Pourhosseingholi A. General family functioning as a predictor of quality of life in parents of children with cancer. J Pediatr Nurs 2019 Jan;44:e2-e8. [doi: 10.1016/j.pedn.2018.08.013] [Medline: 30195919]
51. Qi M, Li P, Moyle W, Weeks B, Jones C. Physical activity, health-related quality of life, and stress among the chinese adult population during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Sep 07;17(18):6494 [FREE Full text] [doi: 10.3390/ijerph17186494] [Medline: 32906604]
52. Luo YH, Li WH, Cheung AT, Ho LL, Xia W, He XL, et al. Relationships between resilience and quality of life in parents of children with cancer. J Health Psychol 2021 Jan 31:1359105321990806. [doi: 10.1177/1359105321990806] [Medline: 33522296]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
QoL: quality of life
RCT: randomized controlled trial

Edited by R Kukafka, G Eysenbach; submitted 01.02.21; peer-reviewed by R Eckhoff, R Chandrasekaran, R Ho, T McCall; comments to author 28.06.21; revised version received 21.07.21; accepted 05.10.21; published 29.11.21.

Please cite as:
Luo Y, Xia W, Cheung AT, Ho LLK, Zhang J, Xie J, Xiao P, Li HCW
Effectiveness of a Mobile Device–Based Resilience Training Program in Reducing Depressive Symptoms and Enhancing Resilience and Quality of Life in Parents of Children With Cancer: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e27639
URL: https://www.jmir.org/2021/11/e27639
doi:10.2196/27639
PMID:34847060

Original Paper

Identifying App-Based Meditation Habits and the Associated Mental Health Benefits: Longitudinal Observational Study

Chad Stecher[1], PhD; Vincent Berardi[2], PhD; Rylan Fowers[1], BSc; Jaclyn Christ[1], MSc; Yunro Chung[1], PhD; Jennifer Huberty[1], PhD

[1]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[2]Department of Psychology, Chapman University, Orange, CA, United States

Corresponding Author:
Chad Stecher, PhD
College of Health Solutions
Arizona State University
500 N 3rd Street
Phoenix, AZ, 85003
United States
Phone: 1 6024960957
Email: chad.stecher@asu.edu

Abstract

Background: Behavioral habits are often initiated by contextual cues that occur at approximately the same time each day; so, it may be possible to identify a reflexive habit based on the temporal similarity of repeated daily behavior. Mobile health tools provide the detailed, longitudinal data necessary for constructing such an indicator of reflexive habits, which can improve our understanding of habit formation and help design more effective mobile health interventions for promoting healthier habits.

Objective: This study aims to use behavioral data from a commercial mindfulness meditation mobile phone app to construct an indicator of reflexive meditation habits based on temporal similarity and estimate the association between temporal similarity and meditation app users' perceived health benefits.

Methods: App-use data from June 2019 to June 2020 were analyzed for 2771 paying subscribers of a meditation mobile phone app, of whom 86.06% (2359/2771) were female, 72.61% (2012/2771) were college educated, 86.29% (2391/2771) were White, and 60.71% (1664/2771) were employed full-time. Participants volunteered to complete a survey assessing their perceived changes in physical and mental health from using the app. Receiver operating characteristic curve analysis was used to evaluate the ability of the temporal similarity measure to predict future behavior, and variable importance statistics from random forest models were used to corroborate these findings. Logistic regression was used to estimate the association between temporal similarity and self-reported physical and mental health benefits.

Results: The temporal similarity of users' daily app use before completing the survey, as measured by the dynamic time warping (DTW) distance between app use on consecutive days, significantly predicted app use at 28 days and at 6 months after the survey, even after controlling for users' demographic and socioeconomic characteristics, total app sessions, duration of app use, and number of days with any app use. In addition, the temporal similarity measure significantly increased in the area under the receiver operating characteristic curve (AUC) for models predicting any future app use in 28 days (AUC=0.868 with DTW and 0.850 without DTW; $P<.001$) and for models predicting any app use in 6 months (AUC=0.821 with DTW and 0.802 without DTW; $P<.001$). Finally, a 1% increase in the temporal similarity of users' daily meditation practice with the app over 6 weeks before the survey was associated with increased odds of reporting mental health improvements, with an odds ratio of 2.94 (95% CI 1.832-6.369).

Conclusions: The temporal similarity of the meditation app use was a significant predictor of future behavior, which suggests that this measure can identify reflexive meditation habits. In addition, temporal similarity was associated with greater perceived mental health benefits, which demonstrates that additional mental health benefits may be derived from forming reflexive meditation habits.

(J Med Internet Res 2021;23(11):e27282) doi:10.2196/27282

KEYWORDS

behavioral habits; habit formation; mindfulness meditation; mental health; mHealth; mobile health; dynamic time warping; mobile phone

Introduction

Background

Practicing healthier daily behaviors would improve many important physical and mental health outcomes for adults in the United States [1-5]. However, even when healthy behaviors are successfully initiated, many people find it difficult to maintain them as long-term habits [6-8] and thus do not attain the corresponding health benefits. One theory of habitual behaviors from psychology and neuroscience contends that habits are unconsciously or reflexively cued by environmental stimuli [9-11]. By repeatedly pairing an environmental or contextual cue with the performance of a desired behavior, our brains routinize the cue-behavior association, reducing the use of deliberative cognitive processes to instigate the daily performance of the behavior [12-14]. Recent research suggests that these contextually cued reflexive responses underlie many of our daily behaviors [9,15], and reflexively instigated habits are a commonly reported behavioral strategy among those who successfully maintain healthy habits, such as daily medication adherence and physical activity [16-20]. However, research has yet to examine mindfulness meditation practices, where the reduction in cognitive effort associated with reflexively instigating meditation [20-22] may enhance the mindfulness experience and increase the corresponding mental health benefits.

An additional limitation to the existing psychology research on habitual behaviors has been the reliance on self-reported measures of habit strength [23]. Habits are theorized to be unconsciously initiated, and as such, individuals should not be able to accurately recall their experience of performing a habitual behavior. Thus, self-reported measures are more likely to capture an individual's perceived self-efficacy or fluency in their behavior [24]. The historical reliance on survey-based habit measures stemmed from a lack of detailed, longitudinal behavioral data necessary for observing daily behavioral patterns. With the recent advent and popularity of mobile health (mHealth) tools that collect precise, high-frequency data on users' daily behaviors, there is a new opportunity for developing more objective indicators of daily habits. These data-driven measures offer the potential to more accurately describe the habit formation process and inform the design of new mHealth interventions that can more successfully promote the formation of healthier habits. As this type of high-frequency daily data are available for many health behaviors, such as physical activity (via wearable fitness trackers), medication adherence (via pill bottles with electronic caps), and mobile phone app–based mindfulness meditation, objective identifiers of reflexive habits will also allow the research on habitual behaviors to be translated across behavioral settings with a higher degree of fidelity.

Goals of This Study

The 2 aims of this study are to construct and test an objective indicator of contextually cued mindfulness meditation habits and to estimate the association between this indicator and improvements in physical and mental health. For this research, we examined detailed observations of mindfulness meditation practices among users of *Calm*, a popular commercial meditation smartphone app. Our novel indicator of contextually cued habits was constructed to capture the temporal similarity of daily app use (ie, using the app at approximately the same time of the day) based on the dynamic time warping (DTW) distance between app use on consecutive days (detailed below). Existing research has shown that most contextually cued habits are performed at approximately the same time and in the same location each day [25,26], which motivates the use of temporal similarity to identify reflexively instigated meditation habits. However, as contextually cued habits are rarely based strictly on time, DTW is used to flexibly measure temporal changes in app-use patterns. For example, a person's daily meditation habit could be cued by finishing lunch or arriving home after work, both of which may occur at slightly different times each day. Accordingly, we apply the DTW measure to capture the broad changes in daily app use start time, duration, or both, while allowing for small daily variations in these dimensions. As a large difference in temporal similarity between 2 days signals that the individual was meditating with the app in a different pattern on those days, we hypothesize that our measure of temporal similarity will indicate when the individual's meditation behavior occurred outside of their usual behavioral context and routines. Admittedly, not all reflexive habits are performed at approximately the same time each day; however, our approach aims to identify most meditation habits that are initiated by temporally similar contextual cues.

In this study, we evaluate our measure of temporal similarity as an indicator of meditation habits by estimating the relative importance of temporal similarity for predicting users' future app use. As contextually cued habits are known to underlie many of our long-term daily behaviors, we hypothesize that our temporal similarity measure will successfully identify many users' reflexive meditation habits and thus significantly predict users' future behavior. We test this hypothesis by comparing the predictive strength of temporal similarity with measures of the frequency and duration of app use [27-31]. Toward the second aim of this study, we then estimate the association between our temporal similarity measure and perceived physical and mental health benefits from using the meditation app. This second set of analyses was exploratory in nature and was designed to investigate the potential relationship between reflexive medication habits and changes in physical and mental health.

This is the first study to offer an objective indicator of reflexive meditation habits based on detailed mHealth data. This measure will allow researchers to better describe the habit formation process and to better target and measure future behavioral

interventions for promoting healthier habits. This study is also the first to examine the potential mental health benefits of forming a mindfulness meditation habit, which provides new directions for future mHealth interventions that aim to improve mental health outcomes.

Methods

Recruitment and Data Description

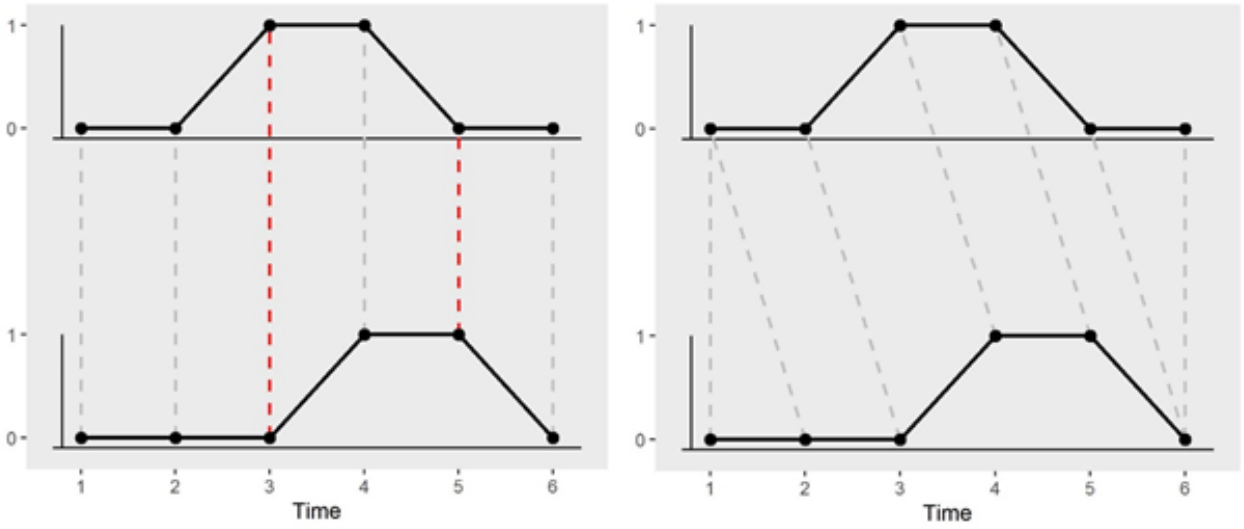
We used longitudinal mindfulness meditation app user data from the commercial app Calm, which had >2 million paying subscribers at the time of data collection. Subscribers in December 2019 who were (1) aged at least 18 years, (2) had a subscription expiration that was at least 2 months away, and (3) had opened at least one email from Calm in the past 90 days were recruited via email to complete a survey on their perceived sleep quality improvements from using the app. These eligibility criteria were used to recruit persistent users of Calm who likely had high intrinsic motivation for meditation, which helps our analyses by reducing the potentially confounding influence of motivation on our ability to identify reflexive meditation habits; the average length of time since first subscribing to Calm was 11.5 months (SD 10.4 months) in this sample. The survey assessment also contained measures of users' socioeconomic status, self-reported app use, physical and mental health status, and perceptions of the physical and mental benefits they experienced after using the app. The survey was approved by the institutional review board of the Arizona State University (STUDY00009725), and the results and additional details on the survey methods and findings have been published elsewhere [32].

Minute-level app use data for the same set of survey respondents were used to develop an objective identifier of habitual app use to predict future use of the app after completion of the survey and to estimate the association with perceived mental health benefits from using the app. The minute-level data set was compiled from all app sessions that were observed 6 months before (June 2019) to 6 months after (June 2020) respondents completed the survey in December 2019. The Calm app offers a wide range of mindfulness content, and a session was recorded when using any one of the following 7 session types: guided or timed meditations, sleep stores, breath training exercise (*breathe*), music, nature sounds (*soundscape*), in-depth audio classes (*masterclass*), and video lessons on mindful movement routines (*body*). For each observed session, the data contained information on the session's start time, duration, type, title of the specific content, and whether the session was completed.

Statistical Analysis

To characterize app use before and after completing the survey, we calculated the number of total sessions used, number of unique sessions used, duration of use, start time, and whether any app use occurred on a given day [27-31]. In addition to these measures, temporal similarity was estimated by converting each day in our sample period into a 1440-minute time series of 0s (when the app was not in use) and 1s (when the app was being used) and then calculating the DTW distance between the time series of app use on consecutive days. As shown in Figure 1, the DTW distance measure is a more flexible calculation than the Euclidean distance between any two 1440-minute time series as it compares the use on one day to use over a wider range of minutes on the other day. Although the Euclidean distance compares use between the exact same minute on each day, Figure 1 shows that the DTW distance is calculated by comparing use on one day with a similar pattern of use that occurred slightly later on the next day. In this way, DTW distance does not penalize users for engaging in the same overall pattern of daily app use that is simply shifted a few minutes earlier or later on a given day. This measure of temporal similarity was calculated using the *dtw* package in R version 4.0.1 (R Foundation for Statistical Computing) and additionally modified to penalize nonuse on consecutive days (see Multimedia Appendix 1 for details). In this way, the final measure of temporal similarity can be interpreted as the average percentage of daily minutes that are inconsistently performed between consecutive days. See Multimedia Appendix 2 for a visualization of the DTW calculation and other attributes found in the Calm app use data.

Figure 1. Visualization of the dynamic time warping (DTW) distance between 2 consecutive days of use, where 0 indicates not using the app and 1 indicates app use during the indicated period. The left panel shows the Euclidean distance between 2 days of equal total use (ie, two time periods with a 1) but the use pattern is shifted in time. As shown on the right panel, DTW compresses or stretches a time series so that the points on one day are mapped to nearby points in the other day that have similar values. The Euclidean distance between these 2 days is 2, whereas the DTW distance is 0. DTW: dynamic time warping.



Each of the objective measures of app use, including frequency, duration, and the temporal similarity measure, were calculated daily for each user and averaged within nonoverlapping 14-day intervals. We then analyzed changes in these measures over multiple intervals to describe how app use patterns changed over time. As some of the users had infrequent app use at the start of our sample period, each user was assigned a unique start date based on the first Monday after at least 5 days of use within a 14-day period. We started each user on a Monday to control for the observed differences in behavior between weekdays and weekends. The objective measures of app use were then defined for all users over 10 nonoverlapping 14-day intervals preceding the date of survey completion. The use of 14-day intervals was chosen so that each interval would contain an equal number of weekdays and weekends, and the results were largely unchanged when these intervals were shortened or widened. For users without 10 nonoverlapping 14-day intervals between their start date and the date of survey completion, additional 14-day intervals were constructed after the survey date so that all users had 10 intervals (140 days) over which the aforementioned objective measures were calculated. These additional postsurvey intervals were not used to analyze the survey-based mental health outcomes (second aim) but were used to test the ability of temporal similarity to predict future app use (first aim).

Finally, these objective measures of app use were calculated based on all observed sessions, as well as separately by session type and session timing, stratified by *Morning* (4 AM to 12 PM), *Evening* (12 PM to 8 PM), or *Night* (8 PM to 4 AM the next day). These measures, along with users' socioeconomic status, were summarized using means and SDs for continuous variables and counts and percentages for dichotomous variables; group differences in dichotomous variables were compared using the *prtest* command in Stata/MP 16.1 (StataCorp) [33].

Toward our first aim of testing temporal similarity as an indicator of reflexive meditation habits, we estimated the predictive ability of temporal similarity on 3 dichotomous measures of future app use: any app use 28 days later, any app use 6 months later, and an identifier for whether the total duration of app use during the 28-day period following the survey was above the median duration during this period. For each of these 3 outcomes, 3 predictive modeling techniques were used. First, logistic regression was used to estimate the associations between each dichotomous outcome and all available objective app-use measures. Second, the same logistic models with and without the DTW-based temporal similarity measures were used to calculate the receiver operating characteristic (ROC) curves and the corresponding areas under the ROC curve (AUCs), a measure of model fit. AUCs from the 2 models (ie, with and without the DTW-based temporal similarity measure) were compared using a chi-square test [34] performed by the *roccomp* command in Stata/MP 16.1. Finally, using the same set of objective app-use measures, we estimated the relative importance of each app-use measure using a random forest algorithm from the *sklearn.ensemble.RandomForestClassier* command in Python 3.8.3 (Python Software Foundation) [35].

To investigate the relationship between reflexive meditation habits and changes in health, we used logistic regression to estimate the association between perceived physical and mental benefits from using the app and the objective app-use measures, including temporal similarity. Physical and mental benefits were measured from survey questions that separately asked whether 7 different physical health conditions (hypertension, high cholesterol, asthma, emphysema, other lung diseases, heart disease, cancer, pain, and arthritis) and 3 different mental health conditions (anxiety, posttraumatic stress disorder, and depression) were *improved* and whether these measures were *very improved*. The dichotomous measure of whether physical

health was *improved* was set equal to 1 if participants responded with an affirmative response (ie, yes, the condition was improved) to any of the physical health conditions and equal to 0 otherwise. Similar procedures were used to construct the dichotomous measure of *improved* mental health and to construct separate dichotomous measures for physical and mental health reported as *very improved*. Logistic regressions estimated the association between these dichotomous measures of physical and mental health improvements and objective app-use measures, controlling for users' socioeconomic status. To better compare model estimates for the different objective app-use measures, a logarithm transformation was applied to each app-use measure so that each association measured the correlation between a 1% increase in the indicated behavioral measure and the odds of experiencing improved health from using the app. Similar log transformations were also applied to the predictive logistic models described above to further improve comparability across the estimated statistical relationships. The logistic regressions were performed by the *logit* command in Stata/MP 16.1

Data Exclusion

Survey respondents that did not have sufficient app use to meet the start date definition detailed above were not included in these analyses (n=2239). However, their irregular app use provides strong evidence that these users did not habitually use the app. Therefore, the results of this study and subsequent discussion serve to describe how *regular* users of the app may benefit from temporally consistent meditation habits, both in their odds of persistent app use and experiencing health benefits.

Results

Sample Characteristics

The sample of Calm users was aged between 21 and 87 years (mean 48.0 years, SD 14.2 years), was primarily female (2359/2771, 86.06%) and White (2391/2771, 86.29%), with a median household income of US $80,000 (mean US $105,927, SD US $86,940.20). Table 1 illustrates the demographic and socioeconomic characteristics, which are largely the same between those with above and equal to or below the median number of meditation sessions used during the 14 days before completing the survey (median=2 sessions). Between both groups, less than two-thirds of users worked full-time. Approximately 43.27% (1199/2771) reported having at least one mental health condition (stress, depression, or anxiety), and approximately 39.01% (1081/2771) reported at least one chronic physical health condition (eg, emphysema or cancer). Household income is presented as a categorical variable in Table 1 to better describe the distribution of income in the sample, which was >US $61,000 for 61.1% (1693/2771) of the participants.

Table 1. Sample demographics by observed use of meditation sessions (N=2771).

Characteristics	Total sample, n (%)	Above median meditation sessions[a] (n=1546), n (%)	Equal or below median meditation sessions[a] (n=1225), n (%)	Difference (percentage points)	*P* value[b]
Age (years)					
18-30	325 (12.23)	186 (12.03)	139 (11.35)	−0.74	.56
31-40	581 (20.97)	317 (20.5)	264 (21.55)	1.05	.50
41-50	607 (21.91)	337 (21.8)	270 (22.04)	0.24	.88
51-60	549 (19.81)	306 (19.79)	243 (19.84)	0.04	.98
61-70	443 (15.99)	254 (16.43)	189 (15.43)	−1.00	.48
71-80	141 (5.09)	73 (4.72)	68 (5.55)	0.83	.32
>81	123 (4.44)	72 (4.66)	51 (4.16)	−0.49	.53
Race					
White	2391 (86.29)	1312 (84.86)	1079 (88.08)	3.22	.01
Asian	78 (2.81)	51 (3.3)	27 (2.2)	−1.09	.08
Native American	6 (0.22)	3 (0.19)	3 (0.24)	0.05	.78
Black	73 (2.63)	45 (2.91)	28 (2.29)	−0.63	.31
Biracial	80 (2.89)	52 (3.36)	28 (2.29)	−1.08	.09
Race other	222 (8.01)	131 (8.47)	91 (7.43)	−1.04	.31
Hispanic	165 (6.05)	98 (6.34)	67 (5.47)	−0.94	.31
Sex (female)	2359 (86.06)	1270 (83.15)	1089 (88.9)	6.66	<.001
Household income (US $)					
<21,000	127 (4.58)	77 (4.98)	50 (4.08)	−1.42	.21
21,000-60,000	469 (16.92)	242 (15.65)	227 (18.53)	3.55	.07
61,000-100,000	521 (18.8)	304 (19.66)	217 (17.71)	−3.34	.10
>100,000	671 (24.21)	368 (23.8)	303 (24.73)	0.57	.79
Employed full-time	1664 (60.71)	910 (59.63)	754 (61.55)	2.42	.20
Education					
Bachelor’s degree	1030 (37.17)	570 (36.87)	460 (37.55)	0.65	.73
Graduate degree	982 (35.44)	559 (36.16)	423 (34.53)	−1.67	.36
Health status					
Mental health condition	1199 (43.27)	680 (43.98)	519 (42.37)	−1.62	.39
Physical health condition	1081 (39.01)	592 (38.29)	489 (39.92)	1.63	.38
Only mental health diagnosis	618 (22.3)	368 (23.8)	250 (20.41)	−3.40	.03

[a]Number of meditation sessions measured over the 14 days before survey completion; median number of meditation sessions over the 14 days before the survey was 2 sessions.

[b]*P* values from proportion tests comparing those with above and those with equal to or below the median number of meditation sessions during the 2 weeks before survey completion.

Most users, approximately 69.87% (1936/2771), reported using the app ≥5 times per week, as shown in Table 2. Importantly, 38% (1053/2771) of users noticed improved mental health status from using the app, whereas 19.38% (537/2771) experienced improved physical health. Table 2 also illustrates the mental health benefits from mindfulness meditation, as the users with above the median number of meditation sessions during the 14 days before completing the survey were 3.5 percentage points (*P*=.05) more likely to experience improved mental health and 8.5 percentage points (*P*<.001) more likely to experience very improved mental health than those with equal to or below the median number of meditation sessions.

Table 2. Self-reported use and health benefits by observed meditation sessions (N=2771).

Characteristics	Total sample, n (%)	Above median meditation sessions[a] (n=1546), n (%)	Equal or below median meditation sessions[a] (n=1225), n (%)	*P* value[b]
Self-reported app use				
1-2 days/week	190 (6.86)	110 (7.12)	80 (6.53)	.55
3-4 days/week	549 (19.81)	280 (18.11)	269 (21.96)	.01
5-7 days/week	1939 (69.97)	1099 (71.09)	840 (68.57)	.15
Use any meditation features	1610 (58.1)	1119 (72.38)	491 (40.08)	<.001
Use any sleep stories features	2282 (82.35)	1165 (75.36)	1117 (91.18)	<.001
Use in mornings	837 (30.21)	689 (44.57)	148 (12.08)	<.001
Use in evenings	809 (29.2)	513 (33.18)	296 (24.16)	<.001
Use at night	2561 (92.42)	1376 (89)	1185 (96.73)	<.001
Try to consistently use at night	1575 (56.84)	790 (51.1)	785 (64.08)	<.001
Perceived health benefits				
Improved mental health	1053 (38)	612 (39.59)	441 (36)	.05
Improved physical health	537 (19.38)	327 (21.15)	210 (17.14)	.008
Improved mental health (only)	771 (27.82)	438 (28.33)	333 (27.18)	.50
Very much improved mental health	482 (17.39)	327 (21.15)	155 (12.65)	<.001
Very much improved physical health	119 (4.29)	75 (4.85)	44 (3.59)	.10
Very much improved mental health (only)	430 (15.52)	295 (19.08)	135 (11.02)	<.001

[a]Number of meditation sessions measured over the 14 days before survey completion; median number of meditation sessions over the 14 days before the survey was 2 sessions.

[b]*P* values from proportion tests comparing those with above and those with equal to or below the median number of meditation sessions during the 2 weeks before completing the survey.

Table 3 presents the mean, SD, and maximum values of the objective measures of app use calculated over the 14 days before survey completion. On average, the likelihood of using any session type on a given day was 55.7%, the likelihood of using meditation features was 28.4%, and the likelihood of using sleep stores was 32% over this period. Approximately 1.14 (SD 2.26) sessions of any type were used per day, with 0.35 (SD 1.11) sessions occurring in the mornings and 0.66 (SD 1.21) sessions occurring in the evenings. Meditation sessions were performed with equal frequency in the mornings—approximately 0.17 (SD 0.36) sessions per day—and the evenings—approximately 0.16 (SD 0.34) sessions per day—whereas sleep stories were largely performed at night (mean 0.37 sessions per day, SD 0.51 sessions per day). A similar pattern was observed for users' average duration in minutes of app use per day by session type, where meditation sessions were used for approximately 2.30 (SD 5.91) minutes in the mornings and 2.84 (SD 6.56) minutes in the evenings, and sleep stories were used for an average of 11.54 (SD 16.19) minutes in the evenings. Finally, the average DTW distance measure across all session types, which incorporated the penalty for nonuse and was standardized by users' average duration of daily use over these 14 days, was equal to 0.539 (SD 0.35). In contrast, the DTW distance measured just among meditation and sleep story features was smaller, at approximately 0.489 (SD 0.44) and 0.474 (SD 0.45), respectively.

Table 3. Objective app use measures over 14 days before survey completion (N=2771).

Characteristics	Value, mean (SD)	Maximum
Any session/day	0.557 (0.37)	1.000
Meditation sessions/day	0.284 (0.36)	1.000
Sleep stories sessions/day	0.320 (0.35)	1.000
Sessions/day in mornings	0.351 (1.11)	48.286
Sessions/day in evenings	0.121 (0.32)	6.786
Sessions/day at night	0.666 (1.21)	45.643
Sessions/day weekdays	0.836 (1.87)	81.643
Sessions/day weekends	0.302 (0.46)	12.286
Sessions/day total	1.139 (2.26)	93.929
Meditation sessions/day in mornings	0.173 (0.36)	4.357
Meditation sessions/day in evenings	0.067 (0.19)	2.571
Meditation sessions/day at night	0.168 (0.34)	3.214
Sleep stories sessions/day in mornings	0.081 (0.25)	2.786
Sleep stories sessions/day in evenings	0.025 (0.12)	2.357
Sleep stories sessions/day at night	0.369 (0.51)	4.929
Duration/day in mornings (minutes)	7.149 (16.82)	352.309
Duration/day in evenings (minutes)	2.137 (5.86)	81.578
Duration/day at night (minutes)	18.043 (25.83)	596.222
Duration/day weekdays (minutes)	19.737 (29.93)	768.200
Duration/day weekends (minutes)	7.592 (11.97)	239.941
Meditation duration/day in mornings (minutes)	2.295 (5.91)	142.167
Meditation duration/day in evenings (minutes)	0.814 (2.36)	35.801
Meditation duration/day at night (minutes)	2.836 (6.56)	83.739
Sleep stories duration/day in mornings (minutes)	2.480 (7.62)	91.848
Sleep stories duration/day in evenings (minutes)	0.750 (3.51)	78.647
Sleep stories duration/day at night (minutes)	11.535 (16.19)	169.736
DTW[a] distance	0.539 (0.35)	3.207
DTW distance (only meditation sessions)	0.489 (0.44)	1.386
DTW distance (only sleep stories)	0.474 (0.45)	1.902

[a]DTW: dynamic time warping.

Predictive Models

The first set of analyses used the objective measure of app use over 140 days to predict users' future app use to test the ability of temporal similarity to identify reflexive habits. Table 4 displays the exponentiated logistic regression coefficients and 95% CIs using the objective app-use measures to predict the 3 dichotomous measures of future app use. The table displays the results for all objective app-use measures calculated over the 6th through the 10th 14-day interval after each user's start date. The app-use measures calculated over intervals 1-5, and the demographic controls displayed in Table 1 were also included in each logistic regression model but were suppressed from Table 4 for ease of presentation.

Table 4. Objective measures in the 10th (2-week) interval predicting future app use (N=2771).[a]

Predictor	Odds ratio (95% CI)					
	Any use 28 days later	Any use 28 days later	Any use 6 months later	Any use 6 months later	High duration in next 28 days	High duration in next 28 days
Days of any use						
Interval 10	3.378[b] (2.32-4.92)	4.282[b] (2.65-6.93)	1.527[c] (1.11-2.10)	1.397 (0.91-2.15)	1.148 (0.66-1.31)	1.676 (0.38-1.80)
Interval 9	1.955[b] (1.37-2.79)	2.159[c] (1.35-3.45)	1.278 (0.91-1.79)	1.197 (0.78-1.84)	1.187 (0.56-1.20)	1.606 (0.32-1.85)
Interval 8	1.119 (0.72-1.74)	1.076 (0.64-1.82)	0.851 (0.56-1.30)	1.089 (0.66-1.81)	1.252 (0.82-1.91)	1.471[d] (1.13-1.97)
Interval 7	0.856 (0.54-1.35)	1.127 (0.65-1.95)	1.165 (0.76-1.78)	1.346 (0.80-2.26)	1.106 (0.72-1.70)	1.195 (0.54-1.71)
Interval 6	1.054 (0.67-1.67)	1.202 (0.71-2.04)	1.115 (0.74-1.68)	1.171 (0.71-1.95)	1.065 (0.71-1.59)	1.203 (0.64-1.63)
Total sessions						
Interval 10	1.347 (0.72-2.52)	1.561 (0.82-2.97)	1.107 (0.58-2.11)	1.024 (0.54-1.95)	1.140 (0.64-2.02)	0.785 (0.37-1.67)
Interval 9	1.508 (0.83-2.73)	1.484 (0.83-2.66)	0.701 (0.39-1.27)	0.757 (0.41-1.41)	0.887 (0.50-1.57)	1.566 (0.78-3.14)
Interval 8	0.899 (0.52-1.55)	0.926 (0.53-1.61)	1.100 (0.62-1.94)	1.019 (0.57-1.82)	1.322 (0.74-2.35)	1.334 (0.61-2.93)
Interval 7	1.127 (0.61-2.07)	1.072 (0.58-1.99)	0.803 (0.45-1.42)	0.835 (0.46-1.53)	1.149 (0.68-1.95)	1.564 (0.74-3.32)
Interval 6	0.809 (0.46-1.41)	0.839 (0.48-1.46)	1.842[d] (1.02-3.34)	1.831 (0.99-3.40)	1.003 (0.56-1.79)	1.043 (0.46-2.34)
Total duration						
Interval 10	0.953 (0.86-1.06)	0.857[d] (0.75-0.97)	1.022 (0.94-1.11)	1.012 (0.92-1.11)	1.108[d] (1.02-1.20)	1.272[c] (1.07-1.51)
Interval 9	0.837[c] (0.74-0.95)	0.796[b] (0.70-0.90)	1.005 (0.92-1.10)	0.984 (0.89-1.08)	0.980 (0.89-1.08)	1.193[d] (1.00-1.42)
Interval 8	1.045 (0.90-1.22)	1.004 (0.86-1.17)	1.020 (0.91-1.15)	1.016 (0.90-1.15)	0.889[d] (0.79-1.00)	0.964 (0.80-1.16)
Interval 7	1.024 (0.89-1.18)	1.022 (0.88-1.18)	1.055 (0.94-1.19)	1.055 (0.93-1.20)	0.912 (0.81-1.02)	1.011 (0.84-1.22)
Interval 6	1.043 (0.90-1.20)	1.064 (0.92-1.23)	0.963 (0.85-1.08)	0.932 (0.82-1.06)	0.964 (0.86-1.08)	1.015 (0.83-1.25)
DTW[e] distance						
Interval 10	—[f]	0.716[d] (0.68-0.92)	—	0.719[c] (0.69-0.88)	—	0.401[c] (0.12-0.81)
Interval 9	—	0.924 (0.32-2.68)	—	0.689 (0.27-1.73)	—	0.896 (0.32-4.30)
Interval 8	—	0.847[c] (0.71-0.93)	—	0.695[d] (0.51-0.81)	—	0.0642[c] (0.01-0.29)
Interval 7	—	0.972 (0.88-6.93)	—	0.941 (0.87-1.17)	—	0.178[d] (0.04-0.77)
Interval 6	—	0.842[d] (0.71-0.92)	—	0.780 (0.30-2.00)	—	0.200 (0.04-1.02)
All-use-measure intervals 5-1	✓[g]	✓	✓	✓	✓	✓
Demographic and SE controls	✓	✓	✓	✓	✓	✓

[a]This table displays the odds ratios (exponentiated coefficients) from separate logistic regression models of each outcome indicated by the column headers on the objective app-use measures indicated by the row labels; 95% CIs are displayed in parentheses. Each objective app-use measure was log transformed to improve the comparability of the estimated relationships, and all models, in addition, included measures of users' demographic and socioeconomic characteristics and the objective app-use measures estimated over intervals 1-5.

[b]$P<.05$.

[c]$P<.01$.

[d]$P<.001$.

[e]DTW: dynamic time warping.

[f]The dynamic time warping distance variables were excluded from one model of each outcome and included in the second model for the same outcome to compare the variable importance results from these 2 approaches.

[g]The indicated variables were also assessed in the model.

Table 4 shows that the number of days of any app use in the 10th 14-day interval (closest interval to the outcomes) was the strongest predictor of the odds of any future use 28 days later (odds ratio [OR] 3.378, 95% CI 2.32-4.92) and the odds of any future use 6 months later (OR 1.527, 95% CI 1.11-2.10). However, this measure was not significantly associated with

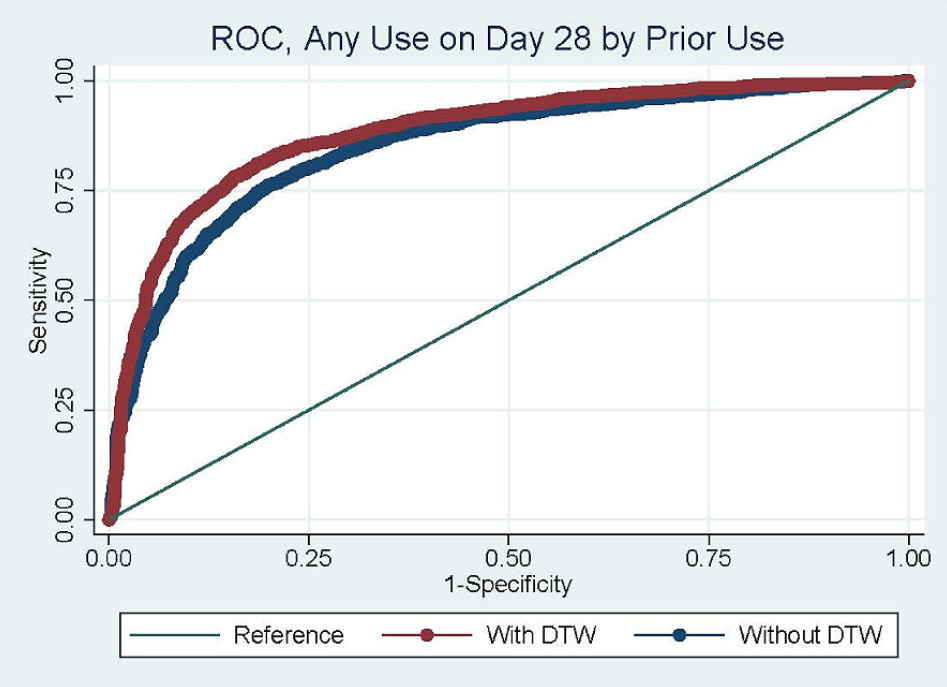
the odds of having a high duration of total app use over the following 28 days. The results also demonstrated that the total number of sessions was a weak predictor of future app use across all 3 outcomes. In addition, total duration was a significant predictor of the odds of having a high duration over the following 28 days (OR 1.108, 95% CI 1.02-1.20). Importantly, the DTW measure of temporal similarity was significantly associated with all 3 outcomes describing future behavior. Specifically, a 1% increase in DTW distance in the 10th 14-day interval (ie, less similar timing of daily app use), which corresponds to a 0.015 SD increase from the mean DTW distance, is associated with an OR of 0.719 (95% CI 0.69-0.88) of any app use 6 months later.

The first panel of Table 5 presents a second statistical test of the predictive value of the temporal similarity measure by comparing the AUC from the same set of logistic regression models shown in Table 5. The AUC without DTW-based temporal similarity measures was 0.818 when predicting any future use 28 days later, and this area was increased by 0.003 (P=.02) when the DTW distance measures were included in the model. To visualize this improvement in prediction accuracy, these 2 ROC curves are displayed in Figure 2. The second panel of Table 5 performs the same comparison between logistic regression models with and without the DTW distance measures, where all objective app-use measures were separately calculated over only the weekdays and weekends in each 14-day interval. These more granularly defined measures of app use improved the overall model fit, as the AUC increased to 0.820 when predicting the odds of any use 28 days later without DTW measures. The AUC increased to 0.828 (P<.001) when including the DTW measures defined over weekdays and weekends. Finally, the third panel in Table 5 displays the results when all objective app-use measures were separately calculated for each of the 7 session types and between weekdays and weekends. When the DTW distance measures were included in these logistic regression models, the AUC was significantly increased for the prediction of any future use 28 days later (P<.001), any use 6 months later (P<.001), and having a high duration of total app use over the next 28 days (P<.001).

Table 5. Area under the receiver operating characteristics curve (AUC) with and without dynamic time warping (DTW) to predict future app use.

Predictor	AUC		
	Any use 28 days later	Any use 6 months later	Above median duration of use over next 28 days
By aggregating all sessions			
Not including DTW	0.818	0.729	0.953
Including DTW	0.821	0.732	0.953
Difference	0.003	0.003	0.000
P value	0.02	0.10	.77
By session timing (weekday, weekend)			
Not including DTW	0.820	0.741	0.950
Including DTW	0.828	0.747	0.956
Difference	0.008	0.008	0.006
P value	<.001	.006	<.001
By session type and timing			
Not including DTW	0.850	0.802	0.958
Including DTW	0.868	0.821	0.963
Difference	0.018	0.019	0.005
P value	<.001	<.001	<.001

Figure 2. The receiver operating characteristic (ROC) curves for logistic regression models that include objective app-use measures calculated over the first 10 14-day intervals from users' start date and demographic controls for predicting the likelihood of any future app use 28 days later. The 2 curves show the models without and with the dynamic time warping (DTW) distance measures, and the area under the ROC curves are equal to 0.850 and 0.868 for the models without and with DTW distance measures, respectively. DTW: dynamic time warping; ROC: receiver operating characteristic.



A second demonstration of the importance of temporal similarity for predicting future app use is presented in Multimedia Appendix 3, which displays the variable importance statistics from random forest models that used the objective app-use measures defined over all session types and timing to predict the same 3 dichotomous measures of future app use presented in Tables 4 and 5. The results show that at least one of the DTW distance measures is among the top 5 predictors of all 3 outcomes, further reinforcing the important association between temporal similarity and behavioral persistence in this setting.

Associated Physical and Mental Benefits

The second set of analyses estimated the association between the objective app-use measures and self-reported mental and physical benefits from using the app. Table 6 displays the exponentiated coefficients from logistic regression models of the dichotomous outcomes indicating users' self-reported *improved* or *very improved* mental or physical health from using the app on the demographic and socioeconomic dimensions presented in Table 1. From Table 6, we can see that older adults were significantly less likely to have reported experiencing mental health improvements but were more likely to have experienced physical health improvements. Specifically, adults aged between 71 and 80 years had an OR 0.239 (95% CI 0.15-0.39) of reporting improved mental health and an OR 2.35 (95% CI 1.15-4.80) of reporting improved physical health relative to users aged under 31 years. These estimated associations highlight the need to include user demographics in subsequent models for predicting perceived health benefits.

Table 6. Self-reported health benefits on user demographics (N=2771).

Demographics	Odds ratio (95% CI)[a]			
	Improved mental health	Improved physical health	Very much improved mental health	Very much improved physical health
Age (years)				
<31	Reference	Reference	Reference	Reference
31-40	0.861 (0.649-1.143)	1.388 (0.817-2.359)	0.815 (0.578-1.150)	2.022 (0.635-6.442)
41-50	0.610[b] (0.458-0.813)	3.586[b] (2.190-5.871)	0.763 (0.539-1.081)	3.409[c] (1.122-10.36)
51-60	0.385[b] (0.286-0.519)	4.729[b] (2.907-7.694)	0.562[d] (0.391-0.808)	6.412[b] (2.236-18.38)
61-70	0.383[b] (0.279-0.526)	6.626[b] (4.040-10.87)	0.613[c] (0.418-0.899)	10.24[b] (3.583-29.29)
71-80	0.239[b] (0.148-0.388)	8.232[b] (4.588-14.77)	0.366[d] (0.191-0.700)	9.842[b] (2.911-33.27)
>81	0.449[d] (0.274-0.737)	2.348[c] (1.150-4.795)	0.712 (0.391-1.296)	3.445 (0.746-15.92)
Sex				
Male	Reference	Reference	Reference	Reference
Female	1.608[b] (1.256-2.060)	0.963 (0.715-1.296)	1.318 (0.965-1.801)	1.490 (0.776-2.861)
Race				
Other	Reference	Reference	Reference	Reference
White	0.936 (0.679-1.291)	0.785 (0.545-1.131)	0.873 (0.601-1.268)	0.879 (0.470-1.644)
Asian	0.336[b] (0.176-0.643)	0.818 (0.377-1.774)	0.397[c] (0.165-0.956)	3.019[c] (1.036-8.803)
Black	0.634 (0.351-1.145)	1.318 (0.706-2.458)	0.737 (0.355-1.531)	2.985[c] (1.189-7.496)
Hispanic	0.731 (0.507-1.055)	1.048 (0.663-1.657)	1.273 (0.848-1.912)	2.801[d] (1.431-5.483)
Log (income)	0.918 (0.829-1.017)	0.888[c] (0.789-0.999)	0.895[c] (0.803-0.998)	0.836[c] (0.707-0.989)
Employment				
Not employed full-time	Reference	Reference	Reference	Reference
Employed full-time	0.847 (0.704-1.019)	0.857 (0.681-1.080)	0.929 (0.737-1.171)	0.933 (0.600-1.451)
Education				
Less than a college degree	Reference	Reference	Reference	Reference
Bachelor's degree	0.787[c] (0.642-0.966)	1.032 (0.800-1.333)	0.736[c] (0.573-0.946)	0.918 (0.580-1.451)
Graduate degree	0.607[b] (0.491-0.751)	0.860 (0.663-1.116)	0.583[b] (0.447-0.760)	0.605 (0.365-1.003)

[a]This table displays the odds ratios (exponentiated coefficients) from separate logistic regression models of each outcome indicated by the column headers on the demographic and socioeconomic characteristics indicated by the row labels.

[b]*P*<.001.

[c]*P*<.05.

[d]*P*<.01.

Table 7 shows the estimated association between improved mental and physical health and the objective measures of app use, conditional on the user demographic and socioeconomic dimensions presented in Table 6. The first panel (model 1) of Table 7 presents the exponentiated coefficients for a model containing the objective app-use measures averaged over the 6 weeks (3×14-day intervals) before survey completion. Conditional on the total number of sessions, total duration of use, and the total number of days with any use, a 1% increase in DTW distance over the 6 weeks before survey completion was associated with an OR 0.34 (95% CI 0.16-0.55) of experiencing improved mental health. The model also estimated a negative relationship between DTW distance and improved physical health; however, this relationship was not statistically significant.

Table 7. Self-reported health benefit on objective app use averaged over the past 6 weeks (N=2771).[a]

Characteristics	Value, odds ratio (95% CI)			
	Improved mental health	Improved physical health	Very much improved mental health	Very much improved physical health
Model 1				
Total number of sessions	1.094 (0.822-1.456)	1.340 (0.965-1.862)	1.718[b] (1.249-2.362)	1.879[c] (1.131-3.121)
Total duration	0.955 (0.874-1.044)	0.929 (0.824-1.048)	0.822[b] (0.733-0.921)	1.006 (0.826-1.225)
Total days with any use	1.244 (0.957-1.618)	1.233 (0.879-1.729)	1.836[b] (1.316-2.562)	0.925 (0.535-1.601)
DTW[d] distance	0.340[e] (0.157-0.546)	0.687 (0.212-1.123)	0.231[e] (0.026-0.487)	0.512[c] (0.315-0.727)
Model 2				
DTW distance; meditation only	0.722[c] (0.495-0.971)	0.601 (0.342-1.022)	0.436[b] (0.277-0.688)	0.475 (0.201- 1.124)
Number of meditation sessions	1.660 (0.876-3.147)	2.180[c] (1.051-4.522)	2.721[e] (1.325-5.587)	0.410 (0.122-1.383)
Duration of meditation sessions	0.985 (0.927-1.048)	0.958 (0.888-1.033)	1.009 (0.937-1.087)	1.090 (0.956-1.242)
DTW distance; sleep stories only	0.715 (0.440-1.430)	0.868 (0.575-1.364)	0.567 (0.311-1.060)	0.914 (0.321-2.601)
Number of sleep story sessions	0.860 (0.541-1.366)	1.369[c] (1.196-2.354)	1.184[c] (1.059-2.026)	1.331[e] (1.189-3.624)
Duration of sleep stories	1.042 (0.993-1.093)	1.021 (0.962-1.085)	1.044 (0.982-1.111)	1.023 (0.910-1.150)
Demographic controls	✓[f]	✓	✓	✓

[a]This table displays the odds ratios (exponentiated coefficients) from separate logistic regression models of each outcome indicated by the column headers on the objective app-use measures indicated by the row labels for each of the 2 predictive models. Each objective app-use measure was log transformed to improve the comparability of the estimated relationships, and all models, in addition, included measures of users' demographic and socioeconomic characteristics.

[b]$P<.001$.

[c]$P<.05$.

[d]DTW: dynamic time warping.

[e]$P<.01$.

[f]Demographic controls were also assessed in the model.

The second panel (model 2) in Table 7 presents the same set of objective app-use measures calculated for the 2 most commonly used session types: meditation and sleep stories. The results show that DTW distance for meditation sessions is a significant predictor of both improved and very improved mental health. Specifically, a 1% increase in DTW for meditation sessions was associated with an OR 0.72 (95% CI 0.50-0.97) of experiencing improved mental health and an OR 0.44 (95% CI 0.27-0.69) of experiencing very improved mental health from using the app. Changes in DTW distance for sleep story sessions were also negatively associated with the odds of experiencing improved (OR 0.72, 95% CI 0.44-1.43) and very improved mental health (OR 0.57, 95% CI 0.31-1.06); however, these relationships were not statistically significant nor were they significant for either measure of physical health improvement. In addition, the DTW distance measure was also a significant predictor of using the app for a greater number of meditation sessions and sleep stories over the next 28 days (Multimedia Appendix 4 and Table S1 of Multimedia Appendix 5).

Discussion

Principal Findings

The 2 aims of this study were to construct and test an objective indicator of reflexive meditation habits and to explore the association between meditation habits and mental health. Toward this first aim, we constructed and provided evidence that our novel measure of temporally similar app use strongly predicted future app use and, thus, is likely to indicate the presence of reflexive meditation habits. Specifically, after controlling for user demographics and common app use metrics, such as frequency and duration of app sessions, the temporal similarity of app use on consecutive days was shown to significantly predict the odds of any future app use and the duration of future use. That is, using the meditation app at roughly the same time each day was associated with greater persistence in meditation app use, which suggests that this measure of temporal similarity was able to indicate the formation of reflexive meditation habits.

We then found that there was a significant association between increased temporal similarity in daily meditation app use and improved mental health from using the app. This finding suggests that the reduced cognitive effort required to instigate meditation that results in a reflexive meditation habit may enable individuals to use more cognitive resources in their mindfulness meditation practice and, thus, experience greater mental health benefits. This interpretation is supported by research describing the impact of attentional resources [36] and mental effort [37-39] on the success of different meditation practices. As the goal of

mindfulness meditation is to direct one's attention to the present moment, these findings indicate that this practice may be easier to perform if meditation is instigated reflexively and without effortful deliberation. However, our study did not investigate the specific mechanism or mechanisms underlying the association between temporal similarity and improved mental health, which is an important area for future research.

Additional Findings

Our descriptive results found that many observable demographic and socioeconomic characteristics were significantly associated with self-reported mental and physical health improvements. Conversely, only the individual's age was significantly associated with measures of the likelihood and duration of future app use (Multimedia Appendix 3). These findings suggest that behavioral habits are equally experienced across demographic characteristics; however, future research assessing the health benefits of mindfulness meditation practices should carefully control for these significant demographic and socioeconomic factors. In addition, the results showed that larger improvements in mental health were reported among individuals who used more than the median number of meditation sessions in the 14 days before the survey, which adds additional evidence for the mental health benefits of mindfulness meditation.

Limitations

Overall, this was a very active sample of app users, with a 55.7% average likelihood of using any app session on a given day over the study period. This is not surprising as the sample consisted of paying subscribers who responded to emails from the app and volunteered to take part in a study to help improve the app; so, the results of this research should be extrapolated to other user types with caution. Although the demographic and socioeconomic minorities of this sample did not display significantly different app-use patterns, the small sample size of these groups limits the statistical power of these comparisons. Future research on mindfulness meditation behaviors should aim to collect data from a wider demographic range of users to better characterize meditation habits and the mental health benefits experienced by all users.

Another limitation was the nature of the data set compiled for these analyses. Specifically, the objective measures of daily app use were compiled for the period around the date when users completed the survey on their perceived mental and physical health benefits. This means that many of the users had been using the app before being included in the sample; so, neither were the analyses able to characterize the initial behavior of app users nor could the results identify the initial period necessary to form a temporally similar meditation habit. In addition, the measure of meditation habits focused on the temporal dimensions of app use as no direct measure was collected on users' behavioral context or environment. Future research that combines the temporal dynamics of daily behavior with additional contextual information could provide a more complete picture of this habit formation process. As temporal data are readily available from a wide range of health apps and other mHealth devices, this study provides a method for analyzing the habit formation process that can be readily applied across these different health behavior settings. Finally, not all reflexive behavioral habits will be performed at approximately the same time each day, and our measure of temporal similarity will only identify the habits that are initiated by temporally similar contextual cues. As past research has found that most habits are instigated by temporally similar contextual cues [25,26], a high degree of temporal similarity is an important indicator of reflexive habits but is not a necessary condition.

Comparison With Prior Work

This study constructed a measure of temporally similar daily meditation app use based on the DTW distance between consecutive days, which adds to the bourgeoning literature using DTW distance to detect temporal patterns in a wide range of behavioral and health data settings [40-45]. One of the most common existing uses of DTW distance has been to categorize patterns of daily health behavior. For example, unhealthy dietary routines were identified by analyzing the variation in DTW distance between participants' eating behaviors over 24-hour periods [43]. DTW distance measures have also been used to identify unhealthy sedentary behaviors [41] and to diagnose hyperactivity disorder from the patterns in children's bodily movements while completing stationary computer tasks [42]. In a clinical setting, DTW-based measures have been used to identify early signs of kidney transplant rejection [40] and to categorize the experience level of surgeons [44]. Recent research has also used DTW-based measures for making predictions from longitudinal health data, such as predicting future glucose levels of patients with type 2 diabetes [45]. In this study, we extended this forecasting approach to the prediction of future health behaviors and showed that a DTW distance measure of temporal similarity in meditation app use on consecutive days significantly predicts an individual's future meditation behavior. This study adds to the nascent literature that uses measures of temporal similarity to characterize health habits [46].

This paper also contributes to the mHealth literature characterizing mobile phone app engagement over time. Declining app use is an important concern and limitation of many app-based health promotion tools, and researchers have found that app engagement durations are becoming increasingly shorter [47]. This is consistent with declining engagement with other mHealth tools over the course of behavioral interventions, which has been observed for self-monitoring technologies, such as physical activity trackers [48] and glucose monitoring [49], as well as adherence to a telemonitoring program for heart failure [50]. This study shows that the temporal similarity of daily meditation app use is an important predictor of continued app use, which suggests that future app-based health promotional tools and interventions should also promote temporal similarity when providing health benefits that require consistent performance over time.

Conclusions

Promoting healthier habits is an important public health objective for improving many health outcomes. However, to date, the study of habit formation has relied on self-reported measures and has yet to use the abundance of new behavioral health data being collected through mHealth devices. The field's limited methods for measuring health habits and understanding of the habit formation process have likely contributed to the

lack of behavioral health interventions that have successfully established persistent behavioral changes.

This study presents a novel objective indicator of reflexive habits derived from detailed, objective behavioral data collected by a mindfulness meditation mobile phone app. Our measure of temporal similarity in meditation app use on consecutive days significantly predicted future app use, even after controlling for users' demographic and socioeconomic characteristics and common measures of app use, such as the frequency and duration of use. Importantly, this temporal similarity measure was also associated with greater odds of experiencing improved mental health from using the app, which suggests that forming a reflexive meditation habit may provide additional mental health benefits. This measure of temporal similarity can be readily applied to other sources of behavioral health data, and future research should build on these findings by investigating the ability of temporal similarity to identify habits in these other behavioral settings. In addition, future research should investigate whether reflexively initiated meditation habits can increase the mental health benefits from mindfulness meditation.

Conflicts of Interest

At the time this paper was written, JH conducted investigator-initiated research and utilized the Calm app to deliver mindfulness meditation interventions. Calm financially supports her research with a postdoctoral scholar and 2 research coordinators (contractors). The roles of these persons are to help JH with her investigator-initiated research using the Calm app. JH consults for Calm as the director of science and director of the scientific advisory board at Calm. Her role is to ensure the quality of Calm's science and she has no specific obligations to the company. She receives no stocks or equity related to the growth or success of the company.

Multimedia Appendix 1
Description of the temporal similarity measure.
[DOCX File , 22 KB - jmir_v23i11e27282_app1.docx]

Multimedia Appendix 2
Visualizations outlining behavioral data from the mindfulness meditation app.
[DOCX File , 493 KB - jmir_v23i11e27282_app2.docx]

Multimedia Appendix 3
Random forest variable importance measures predicting future app use.
[DOCX File , 25 KB - jmir_v23i11e27282_app3.docx]

Multimedia Appendix 4
Logistic regressions of users' future app use on demographic characteristics.
[DOCX File , 24 KB - jmir_v23i11e27282_app4.docx]

Multimedia Appendix 5
Predictive modeling results for alternative measures of users' future app use.
[DOCX File , 30 KB - jmir_v23i11e27282_app5.docx]

References

1. García MC, Bastian B, Rossen LM, Anderson R, Miniño A, Yoon PW, et al. Potentially preventable deaths among the five leading causes of death - United States, 2010 and 2014. MMWR Morb Mortal Wkly Rep 2016 Nov 18;65(45):1245-1255 [FREE Full text] [doi: 10.15585/mmwr.mm6545a1] [Medline: 27855145]
2. Up to 40 percent of annual deaths from each of five leading US causes are preventable. CDC Press Release. 2014. URL: https://www.cdc.gov/media/releases/2014/p0501-preventable-deaths.html [accessed 2019-12-19]
3. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
4. Flett J, Hayne H, Riordan B, Thompson L, Conner T. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10:863-876. [doi: 10.1007/s12671-018-1050-9]
5. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
6. Rodrigues F, Bento T, Cid L, Neiva HP, Teixeira D, Moutão J, et al. Can interpersonal behavior influence the persistence and adherence to physical exercise practice in adults? A systematic review. Front Psychol 2018 Nov 06;9:2141 [FREE Full text] [doi: 10.3389/fpsyg.2018.02141] [Medline: 30459690]

7. Rahimi-Ardabili H, Reynolds R, Vartanian LR, McLeod LV, Zwar N. A systematic review of the efficacy of interventions that aim to increase self-compassion on nutrition habits, eating behaviours, body weight and body image. Mindfulness 2017 Sep 20;9(2):388-400. [doi: 10.1007/s12671-017-0804-0]
8. Zgierska A, Rabago D, Chawla N, Kushner K, Koehler R, Marlatt A. Mindfulness meditation for substance use disorders: a systematic review. Subst Abus 2009;30(4):266-294 [FREE Full text] [doi: 10.1080/08897070903250019] [Medline: 19904664]
9. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev 2007 Oct;114(4):843-863. [doi: 10.1037/0033-295X.114.4.843] [Medline: 17907866]
10. Rothman AJ, Gollwitzer PM, Grant AM, Neal DT, Sheeran P, Wood W. Hale and hearty policies: how psychological science can create and maintain healthy habits. Perspect Psychol Sci 2015 Nov;10(6):701-705. [doi: 10.1177/1745691615598515] [Medline: 26581721]
11. Marteau T, Hollands G, Fletcher P. Changing human behavior to prevent disease: the importance of targeting automatic processes. Science 2012 Sep 21;337(6101):1492-1495. [doi: 10.1126/science.1226918] [Medline: 22997327]
12. Wood W, Neal DT. Healthy through habit: interventions for initiating and maintaining health behavior change. Behav Sci Policy 2016;2(1):71-83. [doi: 10.1353/bsp.2016.0008]
13. Neal D, Wood W, Quinn J. Habits—a repeat performance. Curr Dir Psychol Sci 2006 Aug 01;15(4):198-202. [doi: 10.1111/j.1467-8721.2006.00435.x]
14. Verplanken B. Beyond frequency: habit as mental construct. Br J Soc Psychol 2006 Sep;45(Pt 3):639-656. [doi: 10.1348/014466605X49122] [Medline: 16984725]
15. Rhodes R, Rebar A. Physical activity habit: complexities and controversies. In: The Psychology of Habit. Cham: Springer; Oct 31, 2018:91-109.
16. Phillips LA, Leventhal H, Leventhal EA. Assessing theoretical predictors of long-term medication adherence: patients' treatment-related beliefs, experiential feedback and habit development. Psychol Health 2013;28(10):1135-1151. [doi: 10.1080/08870446.2013.793798] [Medline: 23627524]
17. Brooks TL, Leventhal H, Wolf MS, O'Conor R, Morillo J, Martynenko M, et al. Strategies used by older adults with asthma for adherence to inhaled corticosteroids. J Gen Intern Med 2014 Nov;29(11):1506-1512 [FREE Full text] [doi: 10.1007/s11606-014-2940-8] [Medline: 25092003]
18. Phillips L, Cohen J, Burns E, Abrams J, Renninger S. Self-management of chronic illness: the role of 'habit' versus reflective factors in exercise and medication adherence. J Behav Med 2016 Dec;39(6):1076-1091. [doi: 10.1007/s10865-016-9732-z] [Medline: 26980098]
19. Ellingson L, Lansing J, DeShaw K, Peyer K, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019 Feb 14;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
20. Phillips LA, Gardner B. Habitual exercise instigation (vs. execution) predicts healthy adults' exercise frequency. Health Psychol 2016 Jan;35(1):69-77. [doi: 10.1037/hea0000249] [Medline: 26148187]
21. Gardner B, Phillips L, Judah G. Habitual instigation and habitual execution: definition, measurement, and effects on behaviour frequency. Br J Health Psychol 2016 Sep;21(3):613-630. [doi: 10.1111/bjhp.12189] [Medline: 26991427]
22. Hagger MS. Redefining habits and linking habits with other implicit processes. Psychol Sport Exerc 2020 Jan;46:101606. [doi: 10.1016/j.psychsport.2019.101606]
23. Verplanken B, Orbell S. Reflections on past behavior: a self‐report index of habit strength. J Appl Soc Psychol 2003 Jun;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
24. Hagger MS, Rebar AL, Mullan B, Lipp OV, Chatzisarantis NL. The subjective experience of habit captured by self-report indexes may lead to inaccuracies in the measurement of habitual action. Health Psychol Rev 2014 Dec 18;9(3):296-302. [doi: 10.1080/17437199.2014.959728] [Medline: 25189762]
25. Wood W, Quinn J. Habits and the structure of motivation in everyday life. In: Social Motivation: Conscious and Unconscious Processes. Cambridge: Cambridge University Press; 2004:55-70.
26. Wood W, Quinn J, Kashy D. Habits in everyday life: thought, emotion, and action. J Pers Soc Psychol 2002 Dec;83(6):1281-1297. [doi: 10.1037/0022-3514.83.6.1281] [Medline: 12500811]
27. 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 Jun;94:103187 [FREE Full text] [doi: 10.1016/j.jbi.2019.103187] [Medline: 31026595]
28. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
29. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]

30. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
31. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
32. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia M. 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]
33. Wang D. Confidence intervals for the ratio of two binomial proportions by Koopman's method. Stata Tech Bull 2001;10(58) [FREE Full text]
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988 Sep;44(3):837-845. [doi: 10.2307/2531595] [Medline: 3203132]
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
36. Kaplan S. Meditation, restoration, and the management of mental fatigue. Environ Behav 2001;33(4):480-506 [FREE Full text]
37. Awasthi B. Issues and perspectives in meditation research: in search for a definition. Front Psychol 2013 Jan 10;3:613 [FREE Full text] [doi: 10.3389/fpsyg.2012.00613] [Medline: 23335908]
38. Benson H, Malhotra MS, Goldman RF, Jacobs GD, Hopkins PJ. Three case reports of the metabolic and electroencephalographic changes during advanced Buddhist meditation techniques. Behav Med 1990;16(2):90-95. [doi: 10.1080/08964289.1990.9934596] [Medline: 2194593]
39. Amihai I, Kozhevnikov M. Arousal vs. relaxation: a comparison of the neurophysiological and cognitive correlates of vajrayana and theravada meditative practices. PLoS One 2014 Jul 22;9(7):e102990 [FREE Full text] [doi: 10.1371/journal.pone.0102990] [Medline: 25051268]
40. Fritsche L, Schlaefer A, Budde K, Schroeter K, Neumayer H. Recognition of critical situations from time series of laboratory results by case-based reasoning. J Am Med Inform Assoc 2002;9(5):520-528 [FREE Full text] [doi: 10.1197/jamia.m1013] [Medline: 12223504]
41. Zhang Y, Huang Y, Lu B, Ma Y, Qiu J, Zhao Y, et al. Real-time sitting behavior tracking and analysis for rectification of sitting habits by strain sensor-based flexible data bands. Meas Sci Technol 2020 Feb 06;31(5):055102. [doi: 10.1088/1361-6501/ab63ea]
42. Bautista MA, Hernández-Vela A, Escalera S, Igual L, Pujol O, Moya J, et al. A gesture recognition system for detecting behavioral patterns of ADHD. IEEE Trans Cybern 2016 Jan;46(1):136-147. [doi: 10.1109/TCYB.2015.2396635] [Medline: 26684256]
43. Eicher-Miller HA, Gelfand S, Hwang Y, Delp E, Bhadra A, Guo J. Distance metrics optimized for clustering temporal dietary patterning among U.S. adults. Appetite 2020 Jan 01;144:104451 [FREE Full text] [doi: 10.1016/j.appet.2019.104451] [Medline: 31521771]
44. Forestier G, Lalys F, Riffaud L, Trelhu B, Jannin P. Classification of surgical processes using dynamic time warping. J Biomed Inform 2012 Apr;45(2):255-264 [FREE Full text] [doi: 10.1016/j.jbi.2011.11.002] [Medline: 22120773]
45. Faruqui SH, Du Y, Meka R, Alaeddini A, Li C, Shirinkam S, et al. Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e14452 [FREE Full text] [doi: 10.2196/14452] [Medline: 31682586]
46. Phillips L, Burns E, Leventhal H. Time-of-day differences in treatment-related habit strength and adherence. Ann Behav Med 2021 Mar 20;55(3):280-285 [FREE Full text] [doi: 10.1093/abm/kaaa042] [Medline: 32542355]
47. Vagrani A, Kumar N, Ilavarasan PV. Decline in mobile application life cycle. Procedia Comput Sci 2017;122:957-964. [doi: 10.1016/j.procs.2017.11.460]
48. Whelan ME, Orme MW, Kingsnorth AP, Sherar LB, Denton FL, Esliger DW. Examining the use of glucose and physical activity self-monitoring technologies in individuals at moderate to high risk of developing type 2 diabetes: randomized trial. JMIR Mhealth Uhealth 2019 Oct 28;7(10):e14195 [FREE Full text] [doi: 10.2196/14195] [Medline: 31661077]
49. Tatara N, Arsand E, Skrøvseth SO, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013 Mar 27;1(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2432] [Medline: 25100649]
50. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]

Abbreviations

AUC: area under the receiver operating characteristic curve
DTW: dynamic time warping
mHealth: mobile health
OR: odds ratio
ROC: receiver operating characteristic

Edited by R Kukafka, G Eysenbach; submitted 21.01.21; peer-reviewed by A Phillips, Y Du; comments to author 16.02.21; revised version received 26.02.21; accepted 10.09.21; published 04.11.21.

Please cite as:
Stecher C, Berardi V, Fowers R, Christ J, Chung Y, Huberty J
Identifying App-Based Meditation Habits and the Associated Mental Health Benefits: Longitudinal Observational Study
J Med Internet Res 2021;23(11):e27282
URL: https://www.jmir.org/2021/11/e27282
doi:10.2196/27282
PMID:34734826

Original Paper

Mediating Effects of Stigma and Depressive Symptoms in a Social Media–Based Intervention to Improve Long-term Quality of Life Among People Living With HIV: Secondary Analysis of a Randomized Controlled Trial

Yiran Li[1*], MMSc; Yan Guo[1,2*], PhD; Y Alicia Hong[3], PhD; Chengbo Zeng[4,5], PhD; Yu Zeng[1], MMSc; Hanxi Zhang[6], PhD; Mengting Zhu[7], MS; Jiaying Qiao[8], MMSc; Weiping Cai[9], MD; Linghua Li[9], MD; Cong Liu[9], MSN

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Sun Yat-sen Center for Global Health, Guangzhou, China

[3]Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA, United States

[4]South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[5]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[6]National Center of AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

[7]The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

[8]Department of Vital Statistics, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

[9]Department of Infectious Diseases, Guangzhou Eighth People's Hospital, Guangzhou Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yan Guo, PhD
Department of Medical Statistics
School of Public Health
Sun Yat-sen University
#74 2nd Zhongshan Road
Guangzhou, 510080
China
Phone: 86 020 87334202
Email: guoy8@mail.sysu.edu.cn

Abstract

Background: Mobile health (mHealth) interventions have been shown to effectively improve the quality of life (QOL) among people living with HIV. However, little is known about the long-term effects of mHealth interventions.

Objective: This study aims to explore the intervention mechanisms of a social media–based intervention, Run4Love, on the QOL of people with HIV over across a 9-month follow-up period.

Methods: We recruited people living with HIV who were concurrently experiencing elevated depressive symptoms from an HIV outpatient clinic in South China. A total of 300 eligible participants were randomized either to the intervention group or the control group in a 1:1 ratio after they provided informed consent and completed a baseline survey. The intervention group received a 3-month WeChat-based intervention, comprising cognitive-behavioral stress management (CBSM) courses and physical activity promotion. The control group received a printed brochure on nutrition guidelines in addition to the usual care for HIV treatment. Neither participants nor the research staff were blinded to group assignment. All patients were followed at 3, 6, and 9 months. The primary outcome was depressive symptoms. Structural equation model (SEM) with longitudinal data was conducted to examine the sequential mediating effects of HIV-related stigma and depressive symptoms on the long-term intervention effects on participants' QOL.

Results: About 91.3% (274/300), 88.3% (265/300), and 86.7% (260/300) of all participants completed follow-up surveys at 3, 6, and 9 months, respectively. Results showed that the intervention had significantly improved participants' QOL at 9 months, via complete mediating effects of reduced HIV-related stigma at 3 months and decreased depressive symptoms at 6 months. No adverse events were reported.

Conclusions: These findings underscore the critical roles of HIV-related stigma and depressive symptoms in an mHealth intervention with long-term effects on QOL improvements. We call for targeted mHealth interventions to improve QOL among people living with HIV, especially social media–based interventions that can address HIV-related stigma and alleviate depressive symptoms.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IPR-17012606; https://www.chictr.org.cn/showproj.aspx?proj=21019

(J Med Internet Res 2021;23(11):e27897) doi:10.2196/27897

KEYWORDS

mHealth; HIV; depressive symptoms; quality of life; structural equation model

Introduction

Owing to widely available antiretroviral therapy (ART), living with HIV has become a manageable chronic condition [1]. The 90-90-90 targets that by 2020, 90% of people living with HIV are diagnosed, 90% of those diagnosed receive sustained ART, and 90% of those on treatment are virally suppressed have been unevenly achieved in different contexts [2,3]. In addition, other challenges, such as suboptimal quality of life (QOL), high levels of stigma, and poor mental health among people living with HIV, are still prevalent [4-6]. A great body of evidence has shown that high levels of stigma and depressive symptoms could worsen existing disease conditions and lead to poor QOL [7-9]. A new goal called the fourth "90"—that 90% of people living with HIV have optimal health-related QOL—has been proposed and widely concurred [4]. Emerging research has developed multilevel effective psychosocial, behavioral, and contextualized interventions to improve the QOL among people living with HIV [10].

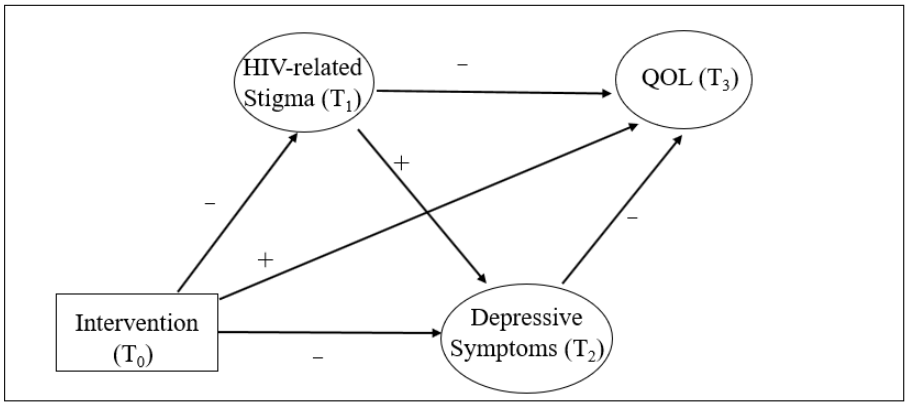
Although psychosocial interventions such as cognitive-behavioral stress management (CBSM) have been proved effective in improving the QOL among people living with HIV, few studies have explored the mechanisms of the effects of such interventions on QOL, especially in the long term [11-14]. Existing studies were often conducted with a pre- and post-design and a short-term follow-up, such as an 8- or 10-week follow-up [15,16]. Few studies examined the long-term intervention mechanisms of QOL improvement [17,18]. Furthermore, most of these interventions were delivered face-to-face in group settings. Given the prevalent HIV-related stigma and discrimination in many countries, including China, a large number of people living with HIV would prefer web-based or mobile-based interventions rather than face-to-face interventions due to privacy concerns.

Mobile health (mHealth) interventions have proved effective in improving QOL among people living with HIV [19-22]; however, to the best of our knowledge, no study has ever explored the intervention mechanisms on improving QOL in mHealth interventions. As mHealth interventions may reach a large number of populations with increased accessibility, convenience, identity protection, and potentially lower cost, such interventions are well suited for marginalized and stigmatized populations, such as people living with HIV [23]. With the increasing number of people living with HIV and wide coverage of smart phones, studies that explore the mechanisms of mHealth interventions on improving the QOL of people living with HIV are much needed, especially those that would examine long-term intervention mechanisms.

Previous studies have suggested that HIV-related stigma and depressive symptoms are important predictors of QOL; however, the mechanisms among these three variables have only been examined in cross-sectional studies, but not in interventional studies [7,24-27]. One cross-sectional study in Ontario, Canada, found that HIV-related stigma was associated with lower QOL, and depressive symptoms partially mediated the association between HIV-related stigma and QOL among African and Caribbean women living with HIV [24]. Another cross-sectional study in Vietnam found that people living with HIV who reported having experienced stigma were more likely to have elevated depressive symptoms and lower QOL [27]. Existing interventional studies have only explored the interventional effects on QOL and potential associated factors using regression analyses. For example, a study found that the improvement of QOL among people living with HIV was positively associated with a reduction in depressive symptoms at the 6-month follow-up in an intervention targeted on depression by regressing the change of QOL on that of depressive symptoms during the 6-month interval [28]. A similar study was also conducted on the association between improvement in QOL and reduction of HIV-related stigma [18]. Due to the limitation of the methodology, existing studies have not yet investigated the mechanisms of intervention effect on QOL.

To investigate the intervention mechanisms, it is important to examine the sequential relationships among influencing factors and the outcome. Structural equation model (SEM) using longitudinal data can provide a better understanding of intervention mechanisms with sequential orders. To fill the gaps of the literature, the current study aimed to explore the intervention mechanisms in the effects of a social media-based intervention, Run4Love, on QOL over 9 months. Using SEM and longitudinal data of 4-time repeated assessments, we examined the sequential relationships among HIV-related stigma, depressive symptoms, and QOL in the mHealth intervention. We hypothesized that the HIV-related stigma and depressive symptoms would play mediating roles in the intervention effects on QOL at the 9-month follow-up. Specifically, the intervention would reduce HIV-related stigma first and subsequently reduce depressive symptoms, which in turn would lead to long-term improvement in QOL among people living with HIV. The hypothesized model is illustrated in Figure 1.

Figure 1. Hypothesized model of intervention, HIV-related stigma, depressive symptoms, and QOL in people living with HIV and depressive symptoms. QOL: Quality of life; T_0: baseline; T_1: 3-month follow-up; T_2: 6-month follow-up; T_3: 9-month follow-up.



Methods

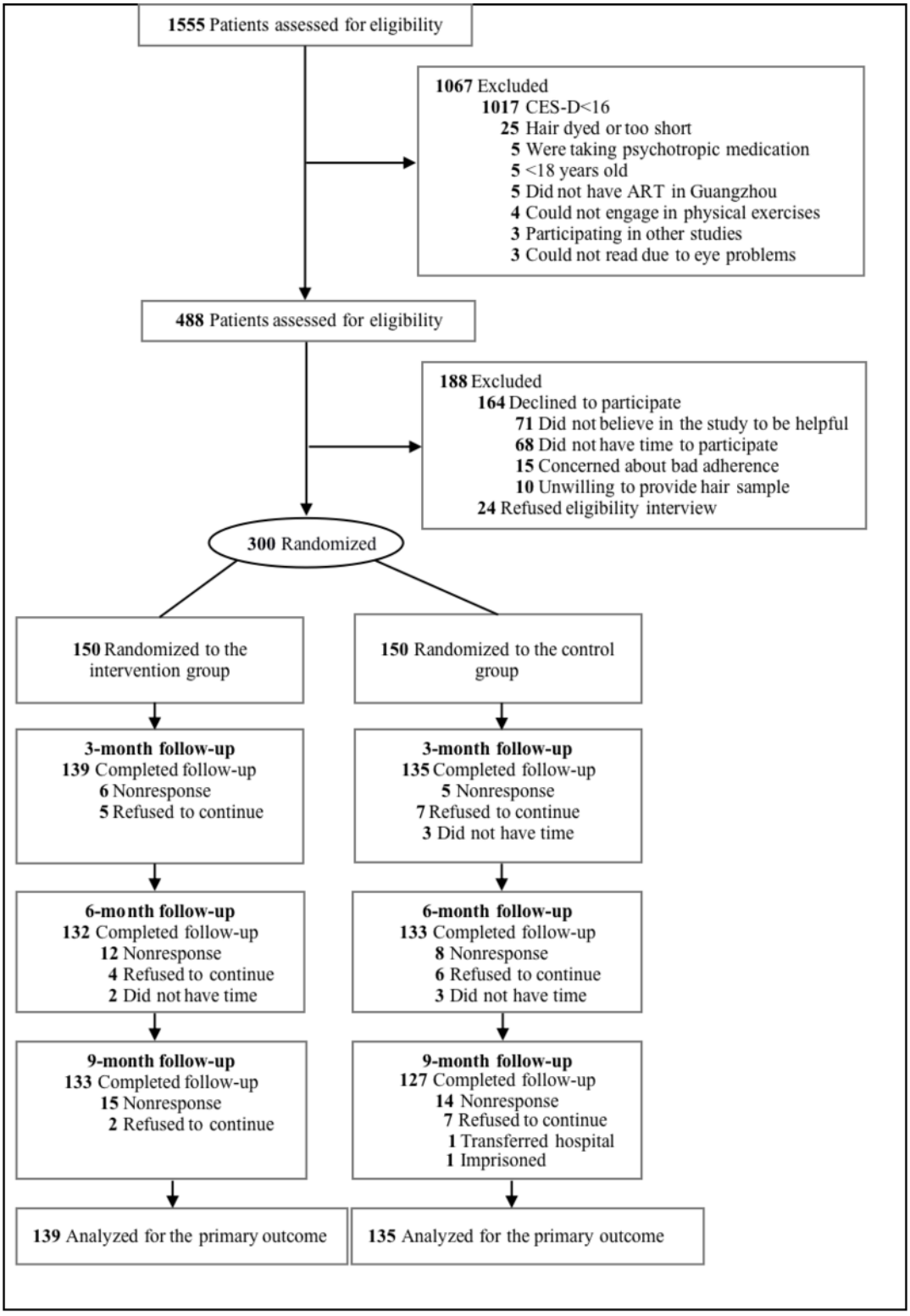
Research Setting

This study used the data from a parallel-design randomized controlled trial (RCT). The CONSORT (Consolidated Standards of Reporting Trials) checklist was used. The study was conducted in Guangzhou, China, between September 2017 and October 2018. A total of 300 people living with HIV and depressive symptoms were recruited by the research staff from a designated hospital for HIV/AIDS treatment in Guangzhou in 2017. Guangzhou is the capital city of Guangdong province, and it is the third largest city in China. The Run4Love RCT protocol was approved by the Institutional Review Board at Sun Yat-Sen University (2018 No. 31) [29].

Participant Eligibility

Participants were recruited if they (1) were aged 18 years or above, (2) tested HIV-seropositive, (3) had elevated depressive symptoms (measured by the Center for Epidemiologic Studies-Depression; CES-D score ≥16), (4) were WeChat users, and (5) were willing to provide hair samples (for the purpose of measuring cortisol as a biomarker of chronic stress). Participants were excluded if they were (1) currently on psychological treatment, (2) unable to complete baseline and follow-up surveys, or (3) unable to engage in the intervention (eg, unable to read or listen to the intervention materials on WeChat or perform physical exercise due to medical or other reasons). Participants who met the eligibility criteria and were willing to participate completed a baseline survey and were randomized to the intervention or the waitlist control group with a 1:1 ratio by the research staff (Figure 2).

Figure 2. Flowchart of participant screening and recruitment. ART: antiretroviral therapy; CES-D: Center for Epidemiologic Studies-Depression.



Run4Love WeChat-Based Intervention

The intervention protocol has been detailed elsewhere [29]. Briefly, the Run4Love WeChat-based intervention comprised two components: adapted CBSM courses and physical activity promotion [29]. The intervention adapted the evidence-based CBSM courses into China's context and included 9 sessions and 3 review sessions with information on meditation, muscle relaxation, and coping skills. The courses were adapted into multimedia formats, including articles, audio clips, and posters. These materials were sent through an enhanced social media platform of WeChat 3 to 5 times a week for 3 months. Physical activity promotion consisted of information on instructions and benefits of regular physical exercise. The enhanced WeChat platform had extended functions, including automatic sending of materials, patient engagement tracking on course completion, and personalized feedback on a weekly basis. In the booster session, a total of 7 most read items were redelivered to the

participants in the intervention group in the 3 months following postintervention. Participants in the control group received a printed brochure on nutrition guidelines in addition to the usual care for HIV treatment.

Sample Size

We calculated the sample size of the RCT based on the primary outcome of depressive symptoms measured by CES-D. Effect size was used to calculate between-group differences. We hypothesized an effect size of 0.4, with α=.05, β=.15, and a dropout rate of 20%, thus arriving at a sample size of 282 for the RCT. Finally, a total of 300 participants were recruited. For SEM, the recommended sample size of 5 observations per estimated parameters was considered [30]. Given a total of 46 estimated parameters in the SEM of our study, the sample size of 300 was sufficient for the analysis.

Randomization and Masking

The research staff randomly assigned the participants to the intervention group or the control group by a computer-generated randomization list with a block size of 4 using SAS software version 9.4 (SAS Institute, Inc). By the nature of the trial design, neither the participants nor the research staff were blinded to the intervention.

Study Outcomes

Overview

Individuals' sociodemographic characteristics, HIV-related stigma, and psychosocial outcomes, including their depressive symptoms and QOL, were included in this study. Sociodemographic characteristics such as age, gender, marital status, sexual orientation, educational level, and household registration (urban and rural) were collected at baseline. Self-reported psychosocial outcomes measured at baseline and at 3-, 6-, and 9-month follow-ups were collected by researchers face-to-face via electronic questionnaires administered using a tablet device.

HIV-Related Stigma

HIV-related stigma was measured by the 14-item HIV Stigma Scale [31,32]. The scale measures 2 dimensions of stigma: perceived stigma and internalized stigma [31]. Perceived stigma comprises 6 items, such as "Most people think that a person who has HIV is dirty"; internalized stigma comprises 8 items, such as "I feel guilty because I have HIV." Each item is rated on a 4-point Likert-type scale (ie, strongly disagree, disagree, agree, and strongly agree). The total score of the scale ranges from 14 to 56, with higher scores indicating higher levels of HIV-related stigma. Total scores ranging from 14 to 28, 29 to 42, and 43 to 56 are considered as low, medium, and high levels of stigma, respectively [33]. In SEM, stigma was measured by the 2 subscales (perceived and internalized stigma). Validity and reliability of the scales have been examined and established for Chinese people living with HIV [32,34]. A good reliability of HIV Stigma Scale was also shown in this study, and the Cronbach α values at baseline, 3-, 6-, and 9-month follow-ups at .92, .95, .95, and .96, respectively.

Depressive Symptoms

Depressive symptoms were assessed using the CES-D scale. The CES-D scale measures 4 dimensions of depressive symptoms, including depressed affect, positive affect, somatic and retarded activity, and interpersonal problems [35,36]. The scale consists of 20 items, such as "I felt depressed" and "I did not feel like eating; my appetite was poor." Items are evaluated using a 4-point Likert scale ranging from 0 (rarely or none of the time) to 3 (most or all the time). A total score of CES-D ranges from 0 to 60, with higher scores indicating a higher level of depressive symptoms. Individuals with CES-D scores no less than 16 are considered as likely having clinical depressive symptoms [37]. The CES-D scale is one of the most widely used self-rated questionnaires on depressive symptoms and has been validated in Chinese populations, including people living with HIV [35,38,39]. In this study, CES-D showed a good reliability, with Cronbach α values reported at .77, .76, .84, and .83 at baseline, 3-, 6-, and 9-month follow-ups, respectively. In SEM, depressive symptoms were measured by the 4 dimensions of the scale.

QOL Assessment

QOL was assessed by World Health Organization Quality of Life HIV–short version (WHO-QOL-HIV BREF), with 31 items measuring the following 6 dimensions: physical, psychological, level of independence, social relationships, environment, and beliefs [40]. Items are rated on a 5-point Likert scale. The total score of WHO-QOL-HIV BREF ranges from 24 to 120, with higher scores indicating better QOL. Validity and reliability of the scale have been examined and established for the Chinese people living with HIV [41-43]. Cronbach α values of WHO-QOL-HIV at baseline, 3-, 6-, and 9-month follow-ups were reported at .84, .91, .94, and .94, respectively.

Data Analysis

First, descriptive statistics were reported on demographic characteristics, baseline HIV-related stigma, depressive symptoms, and QOL. Mean and SD values were used to describe continuous variables with normal distribution, and median and IQR values, for continuous variables with skewed distribution. Frequencies and percentages were used to describe categorical variables. Kolmogorov-Smirnov and Levene tests were used for normality and homogeneity tests.

Second, group differences were examined for the outcome and mediators at 4 assessment points by independent sample *t* test for continuous variables with normal distribution and Wilcoxon rank-sum test for continuous variables with skewed distribution.

Third, bivariate analyses were performed to examine relationships between baseline demographic characteristics and QOL. Independent samples *t* tests were used to examine differences in QOL by different demographic characteristics. Pearson correlation analyses were performed to examine relationships between QOL and continuous demographic variables, such as age. Significance level of potential associations was $P<.10$. Variables significant in bivariate analyses were considered as potential confounders and controlled in the SEM.

Fourth, in SEM analyses, HIV-related stigma, depressive symptoms, and QOL were treated as latent variables and measured by their subscales [44]. SEM allows to control for measurement error of variables; thus, it yields unbiased estimates of latent variables [45,46]. To assess the goodness of fit of the measurement models, confirmatory factor analyses (CFAs) of HIV-related stigma, depressive symptoms, and QOL were performed, respectively. With satisfactory measurement models, the SEM was then conducted to examine the hypothesized mechanisms of long-term intervention effects on participants' QOL.

Finally, a mediation model of SEM was performed to test the hypotheses of sequential mediating effects of HIV-related stigma and depressive symptoms by using longitudinal data. Such models were designed to explore the mechanisms, direct and indirect effects of the intervention, and sequential relationships [47]. Based on the associations between HIV-related stigma, depressive symptoms, and QOL in previous studies, we assigned sequential orders to the SEM variables [24,25,48-50]. Specifically, HIV-related stigma at 3-month follow-up, depressive symptoms at 6-month follow-up, and QOL at 9-month follow-up were estimated using the model. The pathways of intervention→QOL, intervention→depressive symptoms→QOL, intervention→HIV-related stigma→QOL, and intervention→HIV-related stigma→depressive symptoms→QOL were examined separately (dummy coded as 0=control group, 1=intervention group). Baseline QOL and sociodemographic characteristics that were significantly associated with QOL in bivariate analyses were controlled as covariates in the model. The statistical significance in SEM was defined at $P<.05$.

To evaluate the goodness of fit of the model, multiple indicators were used, including the chi-square statistic, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). A smaller chi-square value indicates better model fit. Moreover, CFI ≥0.95, TLI >0.90, RMSEA ≤0.06, and SRMR ≤0.08 indicate good model fit [51,52]. Descriptive statistics, bivariate statistics, and correlation analyses were performed using SAS version 9.4 (SAS Institute, Inc). CFA and SEM were tested by robust maximum likelihood method (estimator=MLR in Mplus) and performed using Mplus version 7.0 [53,54].

Results

Baseline Characteristics

The Run4Love program recruited 300 participants; their baseline characteristics are described in Table 1. Participants' mean age was 28.3 (SD 5.8) years. Most of the participants were male (277/300, 92.3%), nonheterosexual (245/300, 81.7%), not married (262/300, 87.3%), and well educated (high school and above: 182/300, 60.7%). The mean scores of HIV-related stigma, CES-D, and QOL were 37.5 (SD 7.6), 24.1 (SD 6.6), and 77.0 (SD 9.2), respectively, at baseline. A majority (213/300, 71.0%) of the participants had a moderate level of stigma, whereas 20.3% (61/300) had high level and 8.7% (26/300) had low level of stigma.

Table 1. Baseline characteristics and outcomes of study participants by intervention and control group in the Run4Love program (n=300).

Baseline characteristics	Total (n=300)	Intervention group (n=150)	Control group (n=150)	*P* value
Age (years), mean (SD)	28.3 (5.8)	28.0 (5.8)	28.6 (5.9)	.39[a]
Male, n (%)	277 (92.3)	142 (94.7)	135 (90)	.13[b]
Educational level above high school, n (%)	182 (60.7)	98 (65.3)	84 (56)	.10[b]
Homosexual, bisexual, or uncertain, n (%)	245 (81.7)	130 (86.7)	115 (76.7)	.03[b]
Married, n (%)	38 (12.7)	18 (12)	20 (13.3)	.73[b]
Employed, n (%)	251 (83.7)	123 (82)	128 (85.3)	.17[b]
Depressive symptoms, mean (SD)	24.1 (6.6)	23.9 (6.4)	24.3 (6.9)	.68[a]
QOL[c], mean (SD)	77.0 (9.2)	77.4 (9.0)	76.6 (9.4)	.44[a]
HIV-related stigma, mean (SD)	37.5 (7.6)	37.1 (7.7)	38.0 (7.5)	.31[a]
Low, n (%)	26 (8.7)	15 (10)	11 (7.3)	.36[b]
Medium, n (%)	213 (71)	109 (72.7)	104 (69.3)	
High, n (%)	61 (20.3)	26 (17.3)	35 (23.3)	

[a]Based on independent-samples *t* test.

[b]Based on chi-square test.

[c]QOL: quality of life.

The proportion of homosexual, bisexual, or sexual orientation-uncertain participants in the control group was slightly lower than that in the intervention group (115/150, 76.7% vs 130/150, 86.7%; $P=.03$). Other demographic

characteristics and psychosocial outcomes (HIV-related stigma, depressive symptoms, and QOL) were balanced between the two groups at baseline.

Changes in QOL and Mediating Variables Over Time

The proportion of participants who lost to follow-ups was less than 15% over 9 months. Of all participants, 91.3% (274/300; n=139 in the intervention group; n=135 in the control group), 88.3% (265/300; n=132 in the intervention group; n=133 in the control group), and 86.7% (260/300; n=133 in the intervention group; n=127 in the control group) completed the follow-up surveys at 3, 6, and 9 months. The characteristics of the participants who lost to follow-ups were not significantly different from the remaining participants.

As shown in Table 2, the Run4Love social media-based intervention had significant effects on QOL and potential mediating variables over 9 months. Repeated measurements of the outcome variable (QOL) and potential mediators, including HIV-related stigma and depressive symptoms at 4 assessment points for the intervention and control groups, are shown in Table 2. The intervention significantly reduced HIV-related stigma and depressive symptoms, and improved QOL in intervention group participants compared to control group participants at 3-, 6-, and 9-month follow-ups. No adverse events were reported.

Table 2. Repeated measurements of QOL[b] and potential mediators in the Run4Love program.

Variables	Intervention group, mean (SD)	Control group, mean (SD)	Effect size (95% CI)	*t* test (*df*)	*P* value
HIV-related stigma					
Baseline (T_0)	37.10 (7.67)	37.99 (7.54)	N/A[a]	−1.01 (298)	.31
3 months (T_1)	34.28 (9.19)	37.50 (8.27)	0.33 (0.09-0.57)	−3.05 (272)	.003
6 months (T_2)	34.30 (8.52)	37.35 (9.92)	0.27 (0.03-0.51)	−2.69 (263)	.008
9 months (T_3)	33.98 (9.01)	37.79 (9.99)	0.37 (0.12-0.62)	−3.23 (258)	.001
Depressive symptoms					
Baseline (T_0)	23.93 (6.39)	24.25 (6.86)	N/A	−0.42 (298)	.68
3 months (T_1)	17.87 (9.44)	23.85 (10.11)	0.66 (0.42-0.90)	−5.06 (272)	<.001
6 months (T_2)	17.60 (10.06)	24.11 (11.42)	0.63 (0.38-0.88)	−4.93 (263)	<.001
9 months (T_3)	17.86 (10.72)	23.43 (11.45)	0.51 (0.26-0.76)	−4.05 (258)	<.001
QOL[b]					
Baseline (T_0)	77.43 (9.03)	76.59 (9.43)	N/A	0.78 (298)	.44
3 months (T_1)	82.54 (12.03)	76.63 (11.08)	0.55 (0.31-0.79)	4.23 (272)	<.001
6 months (T_2)	83.51 (12.88)	76.32 (12.96)	0.68 (0.43-0.93)	4.53 (263)	<.001
9 months (T_3)	83.48 (13.17)	76.54 (13.34)	0.52 (0.27-0.77)	4.22 (258)	<.001

[a]N/A: not applicable.

[b]QOL: quality of life.

Correlations Between Demographics and QOL

Table 3 shows the correlations between demographic characteristics and QOL at baseline. Bivariate analyses indicated that gender, sexual orientation, education, marital status, and household registration were significantly associated with QOL among people living with HIV and depressive symptoms at baseline. Specifically, those who were male, nonheterosexual, and unmarried; had a higher education level; or had an urban household registration had better QOL at baseline. These factors were included as controlling covariates in the SEM.

Table 3. Bivariate analyses of quality of life (QOL) and demographic characteristics at baseline among people living with HIV and depressive symptoms in the Run4Love program.

Demographics	QOL	*t* value (*df*)	*P* value
Age (years)	N/A[a]	N/A	.76[b]
Gender		1.76 (25)	.09[c]
Male, n (%)	77.33 (9.02)		
Female, n (%)	73.22 (10.9)		
Sexual orientation		−1.95 (72)	.055[c]
Heterosexual, n (%)	74.58 (10.49)		
Homosexual, bisexual, or uncertain, n (%)	77.56 (8.85)		
Education		−3.64 (234)	<.001[c]
Less than or up to high school, n (%)	74.61 (9.51)		
Above high school, n (%)	78.57 (8.72)		
Marital status		2.02 (50)	.049[c]
Unmarried, n (%)	77.40 (9.26)		
Married, n (%)	74.34 (8.65)		
Employment status		−1.19 (274)	.23[c]
Unemployed, n (%)	76.25 (8.67)		
Employed, n (%)	77.52 (9.57)		
Household registration		−2.42 (224)	.02[c]
Rural, n (%)	76.01 (8.97)		
Urban, n (%)	78.67 (9.45)		

[a]N/A: not applicable.

[b]Based on Pearson correlation analysis (r=.02).

[c]Based on independent samples *t* test.

Measurement Model

CFA indicated that the measurement models of HIV-related stigma, depressive symptoms, and QOL yielded good model fit. Indices of the measurement models were reported in Table 4. All factor loadings of each scale were significant at P<.05 level. Results of the standardized factor loadings are shown in Figure 3.

Figure 3. Estimation of the structural equation model. Covariates including gender, sexual orientation, education, marital status, and household registration were controlled in the structural equation model. ** *P*<.01, *** *P*<.001. QOL: Quality of life; T_0: baseline; T_1: 3-month follow-up; T_2: 6-month follow-up; T_3: 9-month follow-up.

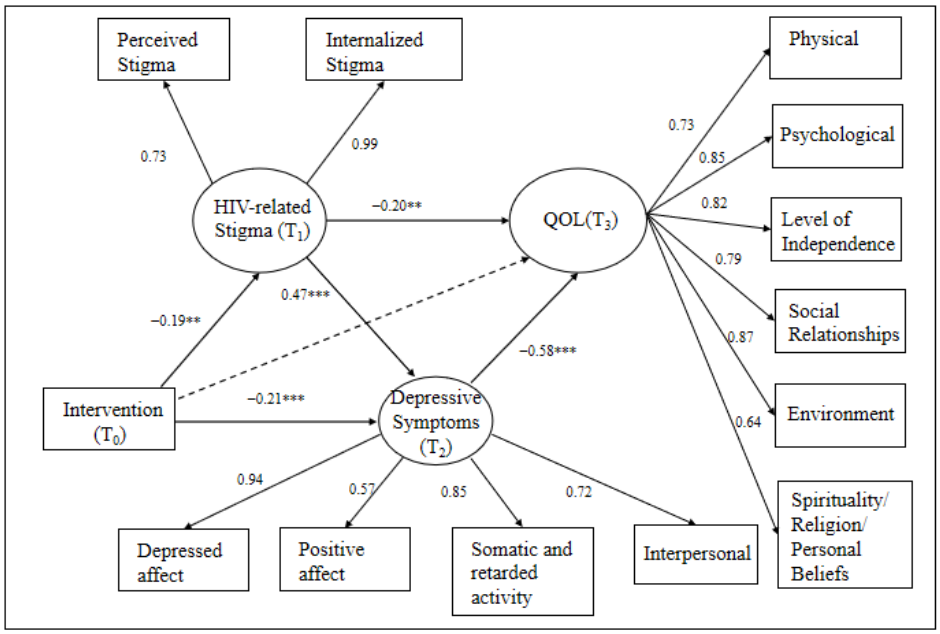


Table 4. Confirmatory factor analyses indices of HIV-related stigma, depressive symptoms, and quality of life.

	Chi-square (*df*)	CFI[a]	TLI[b]	RMSEA[c]	SRMR[d]
HIV-related stigma	0 (0)	1.00	1.00	<0.01	<0.01
Depressive symptoms	3.04 (2)	1.00	1.00	0.04	0.01
QOL[e]	6.47 (8)	1.00	1.00	<0.01	<0.01
Good model fit	N/A	≥0.95	≥0.90	≤0.06	≤0.08

[a]CFI: Comparative Fit Index.

[b]TLI: Tucker-Lewis Index.

[c]RMSEA: root mean square error of approximation.

[d]SRMR: standardized root mean square residual.

[e]QOL: quality of life.

Structural Model

The structural model showed a satisfactory model fit (X^2_{138}=240.4; *P*<.001; CFI=0.93; TLI=0.92; RMSEA=0.06; SRMR=0.06). Demographic characteristics, including gender, sexual orientation, education, marital status, and household registration, were controlled as covariates in the model.

Except for the pathway from intervention to QOL, other pathways were significant, which means that the Run4Love intervention had complete mediation effects on QOL via stigma and depressive symptoms. Standardized regression coefficients for the model are reported in Figure 3 and Table 5. Results indicated that the intervention reduced participants' stigma at the 3-month follow-up (standardized β=−.19, *P*=.001), which was positively associated with depressive symptoms at the 6-month follow-up (standardized β=.47, *P*<.001) and negatively associated with QOL at 9-month follow-up (standardized β=−.20, *P*=.005). The intervention also reduced depressive symptoms (standardized β=−.21, *P*<.001), which consequently resulted in improved QOL at 9 months (standardized β=−.58, *P*<.001). The pathway from depressive symptoms to QOL had the strongest effect size in the mediation model (standardized β=−.58, *P*<.001). There were complete mediation effects of the intervention on QOL via stigma and depressive symptoms. Specifically, the pathway through depressive symptoms alone accounted for the largest mediation effect size of 57.1% (0.12/0.21). The path through stigma→depressive symptoms and stigma alone accounted for 23.8% (0.05/0.21) and 19% (0.04/0.21) of the total effect size on long-term QOL, respectively.

Table 5. Pathway coefficients of the structural equation model.

Pathways	Estimate β	Standardized estimate β	95% CI	SE	*P* value
Intervention→stigma	−2.11	−.19	−3.37 to −0.85	0.64	.001
Intervention→depressive symptoms	−2.05	−.21	−3.11 to −0.99	0.54	<.001
Intervention→QOL[a]	.21	.05	−0.18 to 0.60	0.05	.28
Stigma→depressive symptoms	.41	.47	0.30 to 0.52	0.06	<.001
Stigma→QOL	−.08	−.20	−0.14 to −0.02	0.03	.005
Depressive symptoms→QOL	−.26	−.58	−0.34 to −0.18	0.04	<.001
Total effects from intervention to QOL	1.13	.26	0.60 to 1.66	4.21	<.001
Direct effect	.21	.05	−0.18 to 0.60	0.20	.28
Indirect effects	.92	.21	0.53 to 1.30	0.20	<.001
Intervention→stigma→QOL	.17	.04	0.00 to 0.33	0.08	.048
Intervention→depressive symptoms→QOL	.53	.12	0.23 to 0.82	0.15	<.001
Intervention→stigma→depressive symptoms→QOL	.22	.05	0.07 to 0.38	0.08	.005

[a]QOL: quality of life.

Discussion

Principal Findings

To our knowledge, this study is among the first efforts to explore the potential causal mechanisms of long-term improvement in QOL in an mHealth intervention using longitudinal data. The findings revealed significant mediating roles of stigma and depressive symptoms in improving long-term QOL for the mHealth intervention participants. Specifically, the long-term (9-month follow-up) intervention effect on the improvement of QOL was entirely mediated by the reduction of stigma in the short term (3 months) and reduction of depressive symptoms in the mid term (6 months).

Although several studies have explored the mechanisms of psychosocial interventions, the mechanisms of QOL improvement in the long term remain understudied. For example, studies on mind-body therapies have found that mindfulness might serve as a potentially important mechanism for patients' QOL improvement in such interventions [16,55]. However, literature examining the mechanisms of QOL improvement in psychosocial interventions are scarce, especially in the long term. In addition, methods employed in previous studies such as repeated measures analyses of covariance and latent growth curve model could not identify the sequential and causal relationships when exploring intervention mechanisms [16,17,56]. Our study contributed to the literature by affirming the sequential and potential causal relationships between HIV-related stigma, depressive symptoms, and QOL, which have not been reported in previous studies. Furthermore, only a small number of studies used longitudinal data, of which, to the best of our knowledge, none examined the sequential relationships [16,17,56]. In this study, the use of SEM with longitudinal data allowed us to assign mediators and the outcomes in chronological order, thus shedding light on the potential causal relationships of the intervention mechanisms.

Practical Implications

Given the critical roles of stigma and depressive symptoms in improving long-term QOL, it is important to address stigma and alleviate depressive symptoms in interventions aiming for QOL improvement among people living with HIV. There are several effective ways to reduce HIV-related stigma and to alleviate participants' depressive symptoms as suggested in the literature and shown in this study. First, the intervention content should be evidence-based in reducing HIV-related stigma and alleviating depressive symptoms [5,12]. The intervention materials of the Run4Love program were adapted from the evidence-based and theory-guided CBSM courses, which have been proved effective in reducing stigma and depressive symptoms among people living with HIV [12,57,58]. The core elements of the CBSM courses, such as stress management and coping skills, were preserved and adapted into multimedia formats. The participants in the intervention groups had a moderate level of patient engagement. The cumulative completion rates were 50.6%, 51.5%, and 50.8% at 1, 2, and 3 months, respectively, in the intervention group, which were comparable to other mHealth interventions [59]. The positive relationship between patient engagement and health outcomes has been confirmed in our previous study [59]. Therefore, evidence-based intervention content, rigorous design, and implementation might have contributed to the significant reduction of HIV-related stigma and depressive symptoms in the intervention group.

Second, another effective way to reduce stigma is through social contact; this is one of the most effective strategies to reduce stigma, provide social support, and alleviate depressive symptoms [5,60,61]. Comparatively, mHealth interventions provide varied forms of social contact with better convenience and privacy than face-to-face interventions [62,63]. For example, anonymous virtual communities, such as community message board and web-based forums, serve as easy and secure access for people living with HIV to interact with peers and receive social support [62,63]. Earlier studies suggested that mass media

communication or supportive virtual groups could be incorporated in interventions to mitigate stigma among people living with HIV [62,64]. Such web-based communities and connections are much needed by people living with HIV because of persistent stigma—both perceived and internalized—against HIV and people affected by HIV, as well as the consequent fear of discrimination and social isolation [65,66]. In addition, mHealth interventions allow alternative web connections (eg, human-machine interactions) to incentivize and engage participants. For example, the Run4Love program incorporated automatic weekly feedback on each participant's completion status of the intervention materials, with personalized verbal incentives. Such personalized and patient-centered human-machine interaction is a unique advantage of mHealth interventions. Future mHealth interventions for people living with HIV need to incorporate social contact in various forms as an essential component to address HIV-related stigma and decrease social isolation, so as to reduce depressive symptoms and achieve long-term improvement in people living with HIV's QOL.

Third, since HIV-related stigma is a multidimensional aspect, public policies and coordinated efforts are needed to reduce HIV-related stigma at multiple levels [67,68]. At the societal level, public policies and effective interventions targeting on stigma reduction, HIV testing, ART provision, and social empowerment are needed to address HIV-related stigma among people living with HIV [69]. Community-based support services and educational campaigns should be prioritized to reduce stigma and discrimination [70]. For example, in China, many celebrities, including the First Lady Peng Liyuan, joined the public campaign to reduce discrimination against people living with HIV and promote HIV testing and treatment [71,72]. At the health care level, Voluntary Counseling and Testing program has been instituted at the Centers for Disease Control and Prevention (CDC) and local hospitals [73]. At the individual level, a number of intervention programs focused on stigma reduction, problem-solving skills, and psychological support have been implemented [68].

Limitations

There are some limitations to this study. First, we only focused on HIV-related stigma and depressive symptoms as two important mediators; however, other variables, such as stress, social support, and self-efficacy, might also be potential mediating factors and were not included in this study. Future studies are needed to explore the mediating effects of these factors to better understand the mechanisms of social media-based mHealth interventions. Second, the measurements of stigma, depressive symptoms, and QOL were self-reported and thus might introduce potential recall bias. However, SEM analysis has already considered the measurement errors of these psychosocial variables and therefore provided more reliable model estimations. Third, since the participants were recruited from one hospital in an urban setting and the sample comprised mostly male participants, the generalizability of the study findings to other patients or geographical locations needs to be taken with caution, especially for women and those living in rural areas.

Conclusions

In conclusion, this study is one of the first efforts to examine the potential causal mechanisms of an mHealth intervention in long-term QOL improvement by using longitudinal data and SEM. We found that the long-term intervention effect on QOL improvement was entirely mediated by the reduction of stigma in the short term and reduction of depressive symptoms in the mid term. The findings underscore the critical roles of reducing HIV-related stigma and depressive symptoms in mHealth interventions targeting QOL improvements among people living with HIV. We call for targeted mHealth interventions to improve long-term QOL among people living with HIV, especially social media-based interventions that can reduce HIV-related stigma and alleviate depressive symptoms.

Acknowledgments

This study was supported by a grant from the National Natural Science Foundation of China (grant 71573290) and China Medical Board (grant 17-271). The funders had no role in study design, data collection and analysis, or writing of the manuscript.

Authors' Contributions

YL analyzed the data and drafted the paper. YG and YAH contributed to funding obtaining, study design, and manuscript revision. CZ helped in statistical analysis. YZ, HZ, MZ, JQ contributed to clinical trial and data acquisition. WC, LL, and CL provided administrative, technical, and material support for the clinical trial.

Conflicts of Interest

None declared.

References

1. Hoang NT, Nguyen NTT, Nguyen QN, Bollinger JW, Tran BX, Do NT, et al. Survival outcomes of Vietnamese people with HIV after initiating antiretroviral treatment: role of clinic-related factors. AIDS Behav 2021 May;25(5):1626-1635. [doi: 10.1007/s10461-020-03079-7] [Medline: 33244641]
2. Lazarus JV, Safreed-Harmon K, Barton SE, Costagliola D, Dedes N, Del Amo Valero J, et al. Beyond viral suppression of HIV - the new quality of life frontier. BMC Med 2016 Jun 22;14(1):94 [FREE Full text] [doi: 10.1186/s12916-016-0640-4] [Medline: 27334606]

3. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. UNAIDS. 2014. URL: https://www.unaids.org/en/resources/documents/2017/90-90-90 [accessed 2021-10-14]
4. Webster P. UNAIDS survey aligns with so-called fourth 90 for HIV/AIDS. Lancet 2019 Jun 01;393(10187):2188. [doi: 10.1016/S0140-6736(19)31231-0] [Medline: 34509217]
5. Andersson GZ, Reinius M, Eriksson LE, Svedhem V, Esfahani FM, Deuba K, et al. Stigma reduction interventions in people living with HIV to improve health-related quality of life. Lancet HIV 2020 Feb;7(2):e129-e140 [FREE Full text] [doi: 10.1016/S2352-3018(19)30343-1] [Medline: 31776098]
6. Remien RH, Stirratt MJ, Nguyen N, Robbins RN, Pala AN, Mellins CA. Mental health and HIV/AIDS: the need for an integrated response. AIDS 2019 Jul 15;33(9):1411-1420 [FREE Full text] [doi: 10.1097/QAD.0000000000002227] [Medline: 30950883]
7. Tran BX, Ho RCM, Ho CSH, Latkin CA, Phan HT, Ha GH, et al. Depression among Patients with HIV/AIDS: Research Development and Effective Interventions (GAPRESEARCH). IJERPH 2019 May 19;16(10):1772. [doi: 10.3390/ijerph16101772] [Medline: 31109139]
8. Logie CH, Wang Y, Lacombe-Duncan A, Wagner AC, Kaida A, Conway T, et al. HIV-related stigma, racial discrimination, and gender discrimination: Pathways to physical and mental health-related quality of life among a national cohort of women living with HIV. Prev Med 2018 Feb;107:36-44. [doi: 10.1016/j.ypmed.2017.12.018] [Medline: 29277410]
9. Rao D, Chen WT, Pearson CR, Simoni JM, Fredriksen-Goldsen K, Nelson K, et al. Social support mediates the relationship between HIV stigma and depression/quality of life among people living with HIV in Beijing, China. Int J STD AIDS 2012 Jul;23(7):481-484 [FREE Full text] [doi: 10.1258/ijsa.2009.009428] [Medline: 22844001]
10. Tran BX, Vu GT, Ha GH, Phan HT, Latkin CA, Ho CS, et al. Global mapping of interventions to improve the quality of life of people living with HIV/AIDS: implications for priority settings. AIDS Rev 2020 Feb 12:1-15. [doi: 10.24875/AIDSRev.20000135] [Medline: 32180587]
11. Bhatta DN, Liabsuetrakul T, McNeil EB. Social and behavioral interventions for improving quality of life of HIV infected people receiving antiretroviral therapy: a systematic review and meta-analysis. Health Qual Life Outcomes 2017 Apr 24;15(1):80 [FREE Full text] [doi: 10.1186/s12955-017-0662-4] [Medline: 28438211]
12. Olatunji BO, Mimiaga MJ, O'Cleirigh C, Safren SA. Review of treatment studies of depression in HIV. Top HIV Med 2006;14(3):112-124 [FREE Full text] [Medline: 16946456]
13. Brown JL, Vanable PA. Cognitive-behavioral stress management interventions for persons living with HIV: a review and critique of the literature. Ann Behav Med 2008 Feb;35(1):26-40 [FREE Full text] [doi: 10.1007/s12160-007-9010-y] [Medline: 18347902]
14. Jensen SE, Pereira DB, Whitehead N, Buscher I, McCalla J, Andrasik M, et al. Cognitive-behavioral stress management and psychological well-being in HIV+ racial/ethnic minority women with human papillomavirus. Health Psychol 2013 Feb;32(2):227-230 [FREE Full text] [doi: 10.1037/a0028160] [Medline: 22545977]
15. Penedo FJ, Dahn JR, Molton I, Gonzalez JS, Kinsinger D, Roos BA, et al. Cognitive-behavioral stress management improves stress-management skills and quality of life in men recovering from treatment of prostate carcinoma. Cancer 2004 Jan 01;100(1):192-200 [FREE Full text] [doi: 10.1002/cncr.11894] [Medline: 14692040]
16. Nyklícek I, Kuijpers KF. Effects of mindfulness-based stress reduction intervention on psychological well-being and quality of life: is increased mindfulness indeed the mechanism? Ann Behav Med 2008 Jun;35(3):331-340 [FREE Full text] [doi: 10.1007/s12160-008-9030-2] [Medline: 18535870]
17. Antoni MH, Lechner SC, Kazi A, Wimberly SR, Sifre T, Urcuyo KR, et al. How stress management improves quality of life after treatment for breast cancer. J Consult Clin Psychol 2006 Dec;74(6):1143-1152 [FREE Full text] [doi: 10.1037/0022-006X.74.6.1152] [Medline: 17154743]
18. Van Tam V, Larsson M, Pharris A, Diedrichs B, Nguyen HP, Nguyen CTK, et al. Peer support and improved quality of life among persons living with HIV on antiretroviral treatment: a randomised controlled trial from north-eastern Vietnam. Health Qual Life Outcomes 2012 May 18;10:53 [FREE Full text] [doi: 10.1186/1477-7525-10-53] [Medline: 22606977]
19. Cho H, Iribarren S, Schnall R. Technology-mediated interventions and quality of life for persons living with HIV/AIDS. A systematic review. Appl Clin Inform 2017 Apr 12;8(2):348-368 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0175] [Medline: 28401246]
20. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
21. Lee SB, Valerius J. mHealth interventions to promote anti-retroviral adherence in HIV: narrative review. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e14739 [FREE Full text] [doi: 10.2196/14739] [Medline: 32568720]
22. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management In HIV: a systematic review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
23. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [doi: 10.1007/s11904-007-0029-2] [Medline: 18366952]

24. Logie CH, Ahmed U, Tharao W, Loutfy MR. A structural equation model of factors contributing to quality of life among African and Caribbean women living with HIV in Ontario, Canada. AIDS Res Hum Retroviruses 2017 Mar;33(3):290-297. [doi: 10.1089/AID.2016.0013] [Medline: 27750027]
25. Charles B, Jeyaseelan L, Pandian AK, Sam AE, Thenmozhi M, Jayaseelan V. Association between stigma, depression and quality of life of people living with HIV/AIDS (PLHA) in South India - a community based cross sectional study. BMC Public Health 2012 Jun 21;12:463 [FREE Full text] [doi: 10.1186/1471-2458-12-463] [Medline: 22720691]
26. Rao D, Chen WT, Pearson CR, Simoni JM, Fredriksen-Goldsen K, Nelson K, et al. Social support mediates the relationship between HIV stigma and depression/quality of life among people living with HIV in Beijing, China. Int J STD AIDS 2012 Jul;23(7):481-484 [FREE Full text] [doi: 10.1258/ijsa.2009.009428] [Medline: 22844001]
27. Than PQT, Tran BX, Nguyen CT, Truong NT, Thai TPT, Latkin CA, et al. Stigma against patients with HIV/AIDS in the rapid expansion of antiretroviral treatment in large drug injection-driven HIV epidemics of Vietnam. Harm Reduct J 2019 Jan 17;16(1):6 [FREE Full text] [doi: 10.1186/s12954-019-0277-7] [Medline: 30654814]
28. Bengtson AM, Pence BW, O'Donnell J, Thielman N, Heine A, Zinski A, et al. Improvements in depression and changes in quality of life among HIV-infected adults. AIDS Care 2015;27(1):47-53 [FREE Full text] [doi: 10.1080/09540121.2014.946386] [Medline: 25105320]
29. Guo Y, Hong YA, Qiao J, Xu Z, Zhang H, Zeng C, et al. Run4Love, a mHealth (WeChat-based) intervention to improve mental health of people living with HIV: a randomized controlled trial protocol. BMC Public Health 2018 Jun 26;18(1):793 [FREE Full text] [doi: 10.1186/s12889-018-5693-1] [Medline: 29940921]
30. Bentler PM, Chou C. Practical issues in structural modeling. Sociological Methods & Research 2016 Jun 30;16(1):78-117. [doi: 10.1177/0049124187016001004]
31. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
32. Xiao Z, Li X, Qiao S, Zhou Y, Shen Z, Tang Z. Using communication privacy management theory to examine HIV disclosure to sexual partners/spouses among PLHIV in Guangxi. AIDS Care 2015;27 Suppl 1:73-82 [FREE Full text] [doi: 10.1080/09540121.2015.1055229] [Medline: 26616128]
33. Hong Y, Fang X, Li X, Liu Y, Li M, Tai-Seale T. Self-perceived stigma, depressive symptoms, and suicidal behaviors among female sex workers in China. J Transcult Nurs 2010 Jan 09;21(1):29-34 [FREE Full text] [doi: 10.1177/1043659609349063] [Medline: 19820172]
34. Li Z, Hsieh E, Morano JP, Sheng Y. Exploring HIV-related stigma among HIV-infected men who have sex with men in Beijing, China: a correlation study. AIDS Care 2016 Nov;28(11):1394-1401 [FREE Full text] [doi: 10.1080/09540121.2016.1179713] [Medline: 27137205]
35. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
36. Zhang J, Norvilitis JM. Measuring Chinese psychological well-being with Western developed instruments. J Pers Assess 2002 Dec;79(3):492-511. [doi: 10.1207/S15327752JPA7903_06] [Medline: 12511017]
37. Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. Int J Psychiatry Med 1990;20(3):259-277. [doi: 10.2190/LYKR-7VHP-YJEM-MKM2] [Medline: 2265888]
38. Wang M, Armour C, Wu Y, Ren F, Zhu X, Yao S. Factor structure of the CES-D and measurement invariance across gender in Mainland Chinese adolescents. J Clin Psychol 2013 Sep;69(9):966-979. [doi: 10.1002/jclp.21978] [Medline: 23775279]
39. Zhang Y, Ting RZW, Lam MHB, Lam S, Yeung RO, Nan H, et al. Measuring depression with CES-D in Chinese patients with type 2 diabetes: the validity and its comparison to PHQ-9. BMC Psychiatry 2015 Aug 18;15:198 [FREE Full text] [doi: 10.1186/s12888-015-0580-0] [Medline: 26281832]
40. Hsiung P, Fang C, Wu C, Sheng W, Chen S, Wang J, et al. Validation of the WHOQOL-HIV BREF among HIV-infected patients in Taiwan. AIDS Care 2011 Aug;23(8):1035-1042. [doi: 10.1080/09540121.2010.543881] [Medline: 21500023]
41. Zhu Y, Liu J, Qu B. Psychometric properties of the Chinese version of the WHOQOL-HIV BREF to assess quality of life among people living with HIV/AIDS: a cross-sectional study. BMJ Open 2017 Aug 21;7(8):e016382 [FREE Full text] [doi: 10.1136/bmjopen-2017-016382] [Medline: 28827253]
42. Shan D, Ge Z, Ming S, Wang L, Sante M, He W, et al. Quality of life and related factors among HIV-positive spouses from serodiscordant couples under antiretroviral therapy in Henan Province, China. PLoS One 2011;6(6):e21839 [FREE Full text] [doi: 10.1371/journal.pone.0021839] [Medline: 21738797]
43. Ming Z, Prybylski D, Cheng F, Airawanwat R, Zhu Q, Liu W, et al. Two-year prospective cohort study on quality of life outcomes among people living with HIV after initiation of antiretroviral therapy in Guangxi, China. J Assoc Nurses AIDS Care 2014;25(6):603-613. [doi: 10.1016/j.jana.2014.04.003] [Medline: 24950656]
44. Gribbons BC, Hocevar D. Levels of aggregation in higher level confirmatory factor analysis: application for academic self‐concept. Struct Equ Modeling 2009 Nov 03;5(4):377-390. [doi: 10.1080/10705519809540113]
45. Blanthorne C, Jones-Farmer LA, Almer ED. Why you should consider SEM: A guide to getting started. Bingley: Emerald Group Publishing Limited; Jul 14, 2006.
46. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes 2010 Oct 22;3:267 [FREE Full text] [doi: 10.1186/1756-0500-3-267] [Medline: 20969789]

47. Stanton AL, Luecken LJ, MacKinnon DP, Thompson EH. Mechanisms in psychosocial interventions for adults living with cancer: opportunity for integration of theory, research, and practice. J Consult Clin Psychol 2013 Apr;81(2):318-335. [doi: 10.1037/a0028833] [Medline: 22663900]
48. Rasoolinajad M, Abedinia N, Noorbala AA, Mohraz M, Badie BM, Hamad A, et al. Relationship among HIV-related stigma, mental health and quality of life for HIV-positive patients in Tehran. AIDS Behav 2018 Dec;22(12):3773-3782. [doi: 10.1007/s10461-017-2023-z] [Medline: 29297112]
49. Ekstrand ML, Heylen E, Mazur A, Steward WT, Carpenter C, Yadav K, et al. The role of HIV stigma in ART adherence and quality of life among rural women living with HIV in India. AIDS Behav 2018 Dec 22;22(12):3859-3868 [FREE Full text] [doi: 10.1007/s10461-018-2157-7] [Medline: 29789984]
50. Mahalakshmy T, Premarajan K, Hamide A. Quality of life and its determinants in people living with human immunodeficiency virus infection in puducherry, India. Indian J Community Med 2011 Jul;36(3):203-207 [FREE Full text] [doi: 10.4103/0970-0218.86521] [Medline: 22090674]
51. Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol Methods 1998;3(4):424-453. [doi: 10.1037//1082-989X.3.4.424]
52. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
53. Heck RH, Thomas SL. An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus. Abingdon-on-Thames: Routledge; 2015.
54. Muthén LK, Muthén B. MPlus: statistical analysis with latent variables--User's guide. Los Angeles: Muthe´n & Muthe´n; 2010.
55. Tran BX, Harijanto C, Vu GT, Ho RC. Global mapping of interventions to improve quality of life using mind-body therapies during 1990-2018. Complement Ther Med 2020 Mar;49:102350. [doi: 10.1016/j.ctim.2020.102350] [Medline: 32147055]
56. Zhu M, Cai W, Li L, Guo Y, Monroe-Wise A, Li Y, et al. Mediators of intervention effects on depressive symptoms among people living with HIV: secondary analysis of a mobile health randomized controlled trial using latent growth curve modeling. JMIR Mhealth Uhealth 2019 Nov 15;7(11):e15489 [FREE Full text] [doi: 10.2196/15489] [Medline: 31730042]
57. Heijnders M, Van Der Meij S. The fight against stigma: an overview of stigma-reduction strategies and interventions. Psychol Health Med 2006 Aug;11(3):353-363. [doi: 10.1080/13548500600595327] [Medline: 17130071]
58. Shimotsu S, Horikawa N, Emura R, Ishikawa S, Nagao A, Ogata A, et al. Effectiveness of group cognitive-behavioral therapy in reducing self-stigma in Japanese psychiatric patients. Asian J Psychiatr 2014 Aug;10:39-44. [doi: 10.1016/j.ajp.2014.02.006] [Medline: 25042950]
59. Zeng Y, Guo Y, Li L, Hong YA, Li Y, Zhu M, et al. Relationship between patient engagement and depressive symptoms among people living with HIV in a mobile health intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e20847 [FREE Full text] [doi: 10.2196/20847] [Medline: 33118956]
60. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]
61. Feyissa GT, Lockwood C, Woldie M, Munn Z. Reducing HIV-related stigma and discrimination in healthcare settings: A systematic review of quantitative evidence. PLoS One 2019;14(1):e0211298 [FREE Full text] [doi: 10.1371/journal.pone.0211298] [Medline: 30682131]
62. Flickinger TE, DeBolt C, Xie A, Kosmacki A, Grabowski M, Waldman AL, et al. Addressing stigma through a virtual community for people living with HIV: a mixed methods study of the PositiveLinks mobile health intervention. AIDS Behav 2018 Oct;22(10):3395-3406 [FREE Full text] [doi: 10.1007/s10461-018-2174-6] [Medline: 29882048]
63. Bauermeister JA, Muessig KE, LeGrand S, Flores DD, Choi SK, Dong W, et al. HIV and sexuality stigma reduction through engagement in online forums: results from the HealthMPowerment intervention. AIDS Behav 2019 Mar;23(3):742-752 [FREE Full text] [doi: 10.1007/s10461-018-2256-5] [Medline: 30121727]
64. Tran BX, Phan HT, Latkin CA, Nguyen HLT, Hoang CL, Ho CS, et al. Understanding global HIV stigma and discrimination: are contextual factors sufficiently studied? (GAP Research). Int J Environ Res Public Health 2019 May 29;16(11):1899 [FREE Full text] [doi: 10.3390/ijerph16111899] [Medline: 31146379]
65. Nachega JB, Morroni C, Zuniga JM, Sherer R, Beyrer C, Solomon S, et al. HIV-related stigma, isolation, discrimination, and serostatus disclosure: a global survey of 2035 HIV-infected adults. J Int Assoc Physicians AIDS Care (Chic) 2012;11(3):172-178. [doi: 10.1177/1545109712436723] [Medline: 22431893]
66. Audet CM, McGowan CC, Wallston KA, Kipp AM. Relationship between HIV stigma and self-isolation among people living with HIV in Tennessee. PLoS One 2013;8(8):e69564 [FREE Full text] [doi: 10.1371/journal.pone.0069564] [Medline: 23950897]
67. Rotheram-Borus MJ, Swendeman D, Chovnick G. The past, present, and future of HIV prevention: integrating behavioral, biomedical, and structural intervention strategies for the next generation of HIV prevention. Annu Rev Clin Psychol 2009;5:143-167 [FREE Full text] [doi: 10.1146/annurev.clinpsy.032408.153530] [Medline: 19327028]
68. Hong Y, Li X. HIV/AIDS behavioral interventions in China: a literature review and recommendation for future research. AIDS Behav 2009 Jun;13(3):603-613 [FREE Full text] [doi: 10.1007/s10461-008-9483-0] [Medline: 19015973]

69. Pantelic M, Steinert JI, Park J, Mellors S, Murau F. 'Management of a spoiled identity': systematic review of interventions to address self-stigma among people living with and affected by HIV. BMJ Glob Health 2019;4(2):e001285 [FREE Full text] [doi: 10.1136/bmjgh-2018-001285] [Medline: 30997170]
70. Tran B, Dang A, Truong N, Ha G, Nguyen H, Do H, et al. Depression and quality of life among patients living with HIV/AIDS in the era of universal treatment access in Vietnam. Int J Environ Res Public Health 2018 Dec 17;15(12):2888 [FREE Full text] [doi: 10.3390/ijerph15122888] [Medline: 30562949]
71. Mo Jingxi. Peng Liyuan promotes public health. China Daily. 2018. URL: https://www.chinadaily.com.cn/a/201806/23/WS5b2d8091a3103349141de5a7.html [accessed 2021-10-12]
72. Xu H, Zeng Y, Anderson AF. Chinese NGOs in action against HIV/AIDS. Cell Res 2005;15(11-12):914-918. [doi: 10.1038/sj.cr.7290368] [Medline: 16354569]
73. Xu Z, Ma P, Chu M, Chen Y, Miao J, Xia H, et al. Understanding the role of voluntary counseling and testing (VCT) in HIV prevention in Nantong, China. Biomed Res Int 2020;2020:5740654 [FREE Full text] [doi: 10.1155/2020/5740654] [Medline: 33083474]

Abbreviations

ART: antiretroviral therapy
CBSM: cognitive-behavioral stress management
CDC: Centers for Disease Control and Prevention
CES-D: Center for Epidemiologic Studies-Depression
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
mHealth: mobile health
QOL: quality of life
RMSEA: root mean square error of approximation
SEM: structural equation model
SRMR: standardized root mean square residual
TLI: Tucker-Lewis Index
WHO-QOL-HIV BREF: World Health Organization Quality of Life HIV–Short Version

Edited by R Kukafka; submitted 12.02.21; peer-reviewed by R Ho, M Lotto; comments to author 02.06.21; revised version received 09.06.21; accepted 22.09.21; published 09.11.21.

Please cite as:
Li Y, Guo Y, Hong YA, Zeng C, Zeng Y, Zhang H, Zhu M, Qiao J, Cai W, Li L, Liu C
Mediating Effects of Stigma and Depressive Symptoms in a Social Media–Based Intervention to Improve Long-term Quality of Life Among People Living With HIV: Secondary Analysis of a Randomized Controlled Trial
J Med Internet Res 2021;23(11):e27897
URL: https://www.jmir.org/2021/11/e27897
doi:10.2196/27897
PMID:34751654

Original Paper

Remote Digital Psychiatry for Mobile Mental Health Assessment and Therapy: MindLogger Platform Development Study

Arno Klein[1], PhD; Jon Clucas[1,2*], MIS; Anirudh Krishnakumar[1,3,4*], PhD; Satrajit S Ghosh[5,6], PhD; Wilhelm Van Auken[1], BBA; Benjamin Thonet[1,3], BS; Ihor Sabram[1], MSc IT; Nino Acuna[1], BS; Anisha Keshavan[1,7], PhD; Henry Rossiter[8], HS; Yao Xiao[9], BS; Sergey Semenuta[1], LLM; Alessandra Badioli[1], MA; Kseniia Konishcheva[1,3], MS; Sanu Ann Abraham[5], MS; Lindsay M Alexander[9], MPH; Kathleen R Merikangas[10], PhD; Joel Swendsen[11], PhD; Ariel B Lindner[3], PhD; Michael P Milham[9,12], MD, PhD

[1]MATTER Lab, Child Mind Institute, New York, NY, United States

[2]Computational Neuroimaging Lab, Child Mind Institute, New York, NY, United States

[3]Université de Paris and INSERM U1284 SEED unit, Centre for Research and Interdisciplinarity (CRI), Paris, France

[4]ETH Library Lab, ETH Zurich and Citizen Science Centre, Zurich, Switzerland

[5]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States

[6]Department of Otolaryngology - Head and Neck Surgery, Harvard Medical School, Cambridge, MA, United States

[7]Octave Bioscience, Menlo Park, CA, United States

[8]Computational Engineering, University of Texas at Austin, Austin, TX, United States

[9]Center for the Developing Brain, Child Mind Institute, New York, NY, United States

[10]National Institute of Mental Health, Bethesda, MD, United States

[11]National Center for Scientific Research, University of Bordeaux, EPHE PSL University, Bordeaux, France

[12]Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, United States

[*]these authors contributed equally

Corresponding Author:
Arno Klein, PhD
MATTER Lab
Child Mind Institute
101 East 56th Street
New York, NY, 10022
United States
Phone: 1 347 577 2091
Email: arno@childmind.org

Abstract

Background: Universal access to assessment and treatment of mental health and learning disorders remains a significant and unmet need. There are many people without access to care because of economic, geographic, and cultural barriers, as well as the limited availability of clinical experts who could help advance our understanding and treatment of mental health.

Objective: This study aims to create an open, configurable software platform to build clinical measures, mobile assessments, tasks, and interventions without programming expertise. Specifically, our primary requirements include an administrator interface for creating and scheduling recurring and customized questionnaires where end users receive and respond to scheduled notifications via an iOS or Android app on a mobile device. Such a platform would help relieve overwhelmed health systems and empower remote and disadvantaged subgroups in need of accurate and effective information, assessment, and care. This platform has the potential to advance scientific research by supporting the collection of data with instruments tailored to specific scientific questions from large, distributed, and diverse populations.

Methods: We searched for products that satisfy these requirements. We designed and developed a new software platform called *MindLogger*, which exceeds the requirements. To demonstrate the platform's configurability, we built multiple *applets* (collections of activities) within the MindLogger mobile app and deployed several of them, including a comprehensive set of assessments underway in a large-scale, longitudinal mental health study.

Results: Of the hundreds of products we researched, we found 10 that met our primary requirements with 4 that support end-to-end encryption, 2 that enable restricted access to individual users' data, 1 that provides open-source software, and none that satisfy all three. We compared features related to information presentation and data capture capabilities; privacy and security; and access to the product, code, and data. We successfully built MindLogger mobile and web applications, as well as web browser–based tools for building and editing new applets and for administering them to end users. MindLogger has end-to-end encryption, enables restricted access, is open source, and supports a variety of data collection features. One applet is currently collecting data from children and adolescents in our mental health study, and other applets are in different stages of testing and deployment for use in clinical and research settings.

Conclusions: We demonstrated the flexibility and applicability of the MindLogger platform through its deployment in a large-scale, longitudinal, mobile mental health study and by building a variety of other mental health–related applets. With this release, we encourage a broad range of users to apply the MindLogger platform to create and test applets to advance health care and scientific research. We hope that increasing the availability of applets designed to assess and administer interventions will facilitate access to health care in the general population.

(J Med Internet Res 2021;23(11):e22369) doi:10.2196/22369

KEYWORDS

mental health; mHealth; mobile health; digital health; eHealth; digital psychiatry; digital phenotyping; teletherapy; mobile device; mobile phone; smartphone; ecological momentary assessment; ecological momentary intervention; EMA; EMI; ESM; experience sampling; experience sampling methods

Introduction

Background

This section begins with acknowledging the global burden of mental illness and barriers to care, as well as the increasing need of patients, clinicians, and scientists for digital, mobile, remote mental health care and research. We have then highlighted the advantages of mobile apps, passive monitoring, and experience sampling. To provide context for the rest of the paper, we have described 2 studies: the Healthy Brain Network study and the National Institute of Mental Health (NIMH) Family Study of Affective Spectrum Disorders.

Global Burden of Mental Illness and Barriers to Care

The global burden of mental illnesses is staggering. Epidemiologic studies indicate that 75% of all diagnosable psychiatric disorders begin before the age of 24 years [1] and the lifetime prevalence of a severe disorder among children and adolescents is 21.4%. The most common diagnoses during childhood are anxiety disorders, attention-deficit/hyperactivity disorder, and mood disorders [2]. Despite their high prevalence, only about 50% of children with a mental health disorder receive treatment (Centers for Disease Control and Prevention, NIMH). Adults and children remain untreated, even though effective treatments exist. The World Health Organization has reported large median treatment gaps for alcohol abuse and dependence (78%), generalized anxiety disorder (58%), obsessive-compulsive disorder (57%), depression (56%), dysthymia (56%), panic disorder (56%), bipolar disorder (50%), and schizophrenia (32%) [3,4]. There is a general dearth of mental health resources, especially in low- to middle-income settings [5] and for disadvantaged youth [6], and there are many barriers to care that do not involve finances, insurance, and availability of treatments [6-16]. Simply scheduling an appointment in the United States can be difficult [8], and wait times may take months [8,11]. Additional barriers include concerns about stigma and privacy, as well as exceptional circumstances when telehealth is the only option, such as with the current COVID-19 crisis.

An Increasing Need for Digital, Mobile, Remote Mental Health Care, and Research

People seeking mental health care can use web-based resources to overcome barriers to treatment, such as distance, cost, long waits, inconvenience, stigma, privacy, and problems associated with in-person visits. In recent years, people have been increasingly turning to the internet as a preferred source of knowledge about mental health [7,17,18]. However, the abundance of inaccurate and misleading information on the internet makes it difficult to evaluate the credibility of information sources, the veracity of claims, and the effectiveness of recommendations and therapies. In addition, there are many scenarios in which obtaining data or care without access to the internet or a computer is important. Furthermore, receiving scheduled notifications at any time and place can be critical for a realistic assessment and effective intervention. In the context of these requirements, mobile devices with dedicated apps offer a better way of receiving vetted, curated content and of communicating relevant information without having to sit in front of a computer or find the right person to talk to at the right time.

Clinicians and health organizations, particularly those in the mental health sector, are overwhelmed. General health practitioners, in particular, struggle to meet the needs of their local communities. Better assistance is required to identify mental health issues during checkups to ensure that these concerns are caught early on and that patients benefit from referral to a specialist. Health organizations struggle to provide broader assistance to the vast majority of people around the world who have limited or no access to mental health resources. They need better tools to scale up mental health efforts that reduce or remove face-to-face interactions to inform, assess, and provide therapy. Digital and mobile tools are especially important to reach people who need care and who are in remote

areas, in socioeconomically disadvantaged conditions, or in situations where there is profound stigma surrounding mental illness. Telehealth options that involve video teleconferencing with a health care provider are becoming more prominent and have some evidence of scalability [19]. However, these services require the time and attention of expensive and limited human resources, and therefore do not scale as well as digital, mobile, and remote resources that do not require human interaction. The current COVID-19 crisis has made it starkly apparent that telehealth, with or without human interaction, may at times be the only option for seeking and receiving care. This pandemic is causing a widespread impact on mental health, from anxiety and stress related to helplessness, fear, and uncertainties of the crisis itself to the isolation and loneliness of home confinement and social distancing [20,21]. The need for remotely administered information, assessment, and therapy will only grow as nations have been forced to discover the extensive possibilities of these novel technologies.

Scientists often struggle to acquire relevant mental health and behavioral data when using traditional research paradigms. Study participant recruitment efforts are usually restricted to individuals who can be evaluated in person and provide written consent, which drastically reduces the size and scope of study samples. Smaller sample sizes are less likely to detect effects that exist at the population level, so for data to be relevant, robust, and replicable, it is important to frequently sample from large and diverse populations. Off-the-shelf, web-based tools, such as mobile apps for collecting data, do not meet the needs of most scientists who must configure their data collection tools to reach the right population and address specific questions of scientific and clinical relevance. Even tools that are configurable often require considerable mobile app software engineering expertise that is outside of most laboratories' capabilities. For laboratories attempting to create their own mobile app from scratch, they may turn to outsourcing for its development only to find that it is far more costly than anticipated and requires consistent maintenance and upgrading to keep up with changing software dependencies, operating system versions, and hardware, let alone the changing requirements of the research itself. In addition to the engineering, content, design, and financial challenges of creating mobile apps, there can be considerable logistical and governance challenges involved in dissemination and the incorporation of an electronic consent process.

Appropriately constructed, configured, and vetted mobile apps are currently the most promising avenue for satisfying all of the abovementioned requirements, as mobile devices with internet access are rapidly becoming widely available to diverse populations around the world. It is currently estimated that 60% of the world's population uses the internet and that 93% of users gain internet access through mobile devices [22]. They provide a convenient way to passively or actively collect and present information that can be relevant to the natural variations and social and environmental contexts of people's lives. Passive methods usually involve sensors carried or worn on some part of the body and can collect objective, real-world data about participants to monitor motor activity, sleep, heart rate, cognition, behaviors, mood, and physiological states, as well as detect important outcomes such as medication response [23]. Some recent examples of sensor-based technologies that track mental health include phone apps, such as Northwestern University's Center for Behavioral Intervention Technologies' Purple [24,25] and Intellicare [26,27] platforms, Harvard University's Beiwe platform [28,29], and apps such as mPower [30,31] built on top of Apple's ResearchKit for iOS [32] and ResearchStack for Android [33]. The Beth Israel Deaconess Medical Center maintains an extensive database of apps that claim to respond to mental health needs [34] and provides a protocol for their evaluation [35,36]. There is an unmet need for a free, open, configurable platform to translate pencil-and-paper assessment instruments, cognitive tasks, and therapies into attractive, engaging, and effective digital, mobile, and remote tools that exceed current standards of privacy, security, and accessibility.

One important limitation in the interpretation of data acquired through passive monitoring is the lack of contextual information on variables that may influence sleep, activity, or mood changes inferred from speech, texting, GPS location, or other interactions with mobile devices. Although studies have combined subjective symptom ratings with passive monitoring, retrospective reporting over the past week or month is common among many traditional clinical questionnaires and may miss important real-time associations between subjective and objective data. Gaining insight into the directional associations between events and psychological states and their association with sleep and physical activity can be enhanced through the administration of tools that simultaneously capture descriptions of symptoms of mood, cognition, and other subjective experiences, which remain the core components of psychiatric disorder criteria. These active methods of recording individuals' internal states and experiences include explicit self-reports that may range from occasional and detailed survey instruments to more frequent, brief, and in-the-moment questionnaires. Repeated assessment of people in real time is best known as *experience sampling* [37] and in the context of medicine and physiological or event-related data capture has been referred to as ecological momentary assessment (EMA) [38]. As pointed out by Marije aan het Rot et al [39] in their excellent review of experience sampling methods and EMA [39], the 2 terms are increasingly used together, so we will simply refer to them as *experience sampling*.

As described by Trull and Ebner-Priemer [40], experience sampling offers major benefits over traditional clinical assessments, including the reduction of retrospective bias, real-time tracking of dynamic processes, simultaneous integration of multi-level data (eg, biological and psychological), characterization of context-specific relationships of behaviors and symptoms, inclusion of feedback, and enhanced generalizability of results. Experience sampling has been shown to be highly feasible and valid in the assessment of diverse categories of mental illness, including patients with mood disorders, anxiety disorders, substance use disorders, and psychosis [41-45] as well as in the assessment of transdiagnostic mental health issues such as suicidal ideation [46]. More recently, experience sampling has been used to assess cognitive functions [47]. Batteries of cognitive tasks, such as the NIH Toolbox [48,49], ACE [50], and Cambridge Cognition [51], are

primarily used for research and are not currently intended for clinical practice; they are for the most part commercial, proprietary, and permit limited (if any) configuration options for presenting and collecting data. Although mental health professionals may not adopt new technologies for clinical assessment more readily than traditional tools [52], their use in clinical practice may be encouraged by increasing their personalization and integration into standard care [53].

The Healthy Brain Network Study and NIMH Family Study of Affective Spectrum Disorders

The Child Mind Institute's Healthy Brain Network study [54] is an ongoing initiative focused on creating and sharing a biobank of data from 10,000 New York area participants (age group 5-21 years). The biobank houses data on psychiatric, behavioral, cognitive, and lifestyle phenotypes, as well as multimodal brain imaging (resting and naturalistic viewing functional magnetic resonance imaging (MRI), diffusion MRI, and morphometric MRI), electroencephalography, eye tracking, voice and video recording, genetics, and actigraphy.

The NIMH Family Study of Affective Spectrum Disorders was a large, community-based, controlled family study [55] that collected assessment data 4 times per day for 2 weeks from phones provided to participants of the study. The assessments included questions about daily life experiences and behaviors at the moment of acquisition (current location, social company, performance of specific behaviors, and mood states) and since the previous assessment or since waking up (experience of daily events and event negativity, food intake, substance use, experience of headache, and specific symptoms). Assessments at the beginning of the day also included questions about duration, quality, and problems with sleep, and at the end of the day included ratings of the stressfulness of the day, food craving for the day, and specific physical symptoms (gastrointestinal symptoms and muscle pain). Response options included Likert scales for dimensional constructs (such as mood or event negativity) and checklists for multiple responses (such as for food consumed) or single responses (such as current physical location).

The data from the NIMH Family Study that evaluated the association between daily events and emotional experience yielded important differences in patterns of reactivity among the major subtypes of mood disorders, including bipolar I disorder, bipolar II disorder, major depression, anxiety disorders without a mood disorder, and controls [56]. These findings demonstrate how experience sampling is a particularly well-adapted tool for assessing affective dynamics as well as emotional reactivity following daily life events. The value of combined passive and active monitoring in this study further showed bidirectional associations between energy, motor activity, and sleep, and unidirectional associations between activity and mood, suggesting that increased activity could be used as an intervention for depression [57]. Using the novel analytic approach of fragmentation to test the stability and instability of emotional states in this study showed greater instability of energy and attention in people with a history of bipolar I disorder, whereas those with bipolar II disorder or major depression exhibited greater fragmentation of mood and anxiety [58].

Outline of the Paper

In this paper, we (1) review customizable, mobile, experience sampling products for configurable data collection and content delivery, (2) summarize the motivation for and development of a new mobile platform called MindLogger, and (3) describe MindLogger applets, including an initial use case that applies a MindLogger version of the NIMH Family Study's app as part of the Healthy Brain Network study.

Methods

Overview

In this section, we have discussed our criteria for selecting and evaluating customizable, mobile, experience sampling products; provided an overview of the development of the MindLogger platform for experience sampling and interventions; and presented an example app of MindLogger in a large-scale mental health study. The *Results* section follows up on each of these, presenting the results of the product review, the current state of the MindLogger platform, and example apps of MindLogger.

Criteria for Reviewing Customizable, Mobile, and Experience Sampling Products

In this review, we did not consider mobile apps limited to specific assessments, cognitive tasks, or therapies, but rather examined platforms that enable the creation and distribution of such apps. We wanted to find products that have an administrator interface for creating and scheduling recurring, customized questionnaires, where users receive and respond to scheduled notifications on a mobile device.

The detailed outline of the protocol we followed has been provided in Multimedia Appendix 1, which is briefly summarized in the paper. Over the last 3 years, clinical and research collaborators and colleagues have helped us gather information about products with desired characteristics. To extend this search, we conducted 3 queries in Google's search engine (without quotation marks or Boolean operators): (1) *digital electronic data capture systems*, to broadly identify any electronic tools for capturing data; (2) *mobile phone software sensor data collection*, to identify mobile data collection software that may involve sensors; and (3) *alternative to Qualtrics*, to identify alternatives to one of the most prominent products that enables web-based customization of surveys (although Qualtrics itself does not currently have a mobile app). We visited the websites of the first 20 search results from each query (not including advertisements) and identified the candidate products. For example, if a website listed the *Top Ten Apps for...*, we would include these 10 apps in our initial set of candidates. We then filtered this set using the following inclusion criteria: the candidate product had to (1) be in current use, (2) have (Android and iOS) mobile apps, and (3) have an administrative user interface for creating and scheduling times and days for recurring, customized questionnaires, where end users receive and respond to scheduled notifications via an iOS or Android app on a mobile device. Where there was any ambiguity, we contacted the company or organization to clarify,

scheduled a web-based demonstration, and requested a free trial to explore the product. We excluded products that do not currently fulfill the above requirements or for which we could not receive a demonstration or trial without a legal agreement. We also excluded products that require SMS text messaging, email, or other modes of communication outside of their mobile app to send and receive notifications.

Because the type of notification is important in experience sampling applications, we identified which products can deliver local operating system notifications or push notifications, where an end user receives notifications in their mobile device's notification bar at scheduled times, and a tap on a notification takes them to their scheduled activity within the mobile app. Local operating system notifications do not require an internet connection at the time that the notification is to be received, whereas push notifications do, and both of these are distinct from simple in-app notifications, which require the end user to use the app to see their notifications.

We collected additional information from product websites, and via teleconferences and web-based demonstrations with the product creators, and from free trials to determine the degree to which each is (1) customizable in its information presentation and data capture capabilities, (2) private and secure, and (3) accessible (easy to use, economical, and open source). Customization is important to ensure that content, language, and presentation can be adapted and updated to meet the specific needs of a given population of intended end users, and to expand the scope of possible dimensions to assess, analyses to perform, and inferences to make. Privacy and security is a central concern, especially as we conduct research involving a triply vulnerable population of end users: (1) child and adolescent (2) patients with (3) mental health and learning disorders. Data access, encryption, and deletion capabilities are the primary considerations in this domain. *Accessible* can mean many things; here we refer to whether and how the administrator can access the product, its source code, or its data. The degree to which a product is affordable will often determine the degree to which it is adopted. Possibly, the most stringent accessibility criterion one could have is for the software to be open source. Open-source software is important because the product does not live or die with a given company, provides an opportunity for anyone to build on and improve the software, and it is open to greater scrutiny to ensure the quality of the software, accuracy of any claims made about the software, and transparency of clinical and scientific practices that use the software. Each of these characteristics offers a competitive advantage; however, they should be considered together. For example, a company offering a free product can store or transmit user data insecurely.

Development of the MindLogger Platform

MindLogger [59] is intended as an *open ecosystem to create, edit, share, and administer mobile or web applets for data collection and content delivery*. We use the term *applet* to refer to a customized collection of activities within the MindLogger app administered to target end users. Our focus with MindLogger is to easily create and edit digital mental health assessments and interventions and administer them to users via mobile or web applets. The 3 key innovations we sought to accomplish with MindLogger were *customizability* of content, response options, and appearance; an extensive library of applets built using *open standards* (open, reusable parts defined by an open protocol), and distribution as a *single app* that appears differently to different user groups.

MindLogger's development began in 2017, with a focus on mental health assessments. After the significant development of an early prototype for mental health research and clinical colleagues at the Child Mind Institute and the Child Mind Medical Practice in New York City, we revisited development with a greater emphasis on human-centered design [60,61]. We included a variety of key stakeholders to ensure that their needs would be met by the platform: clinicians (psychiatrists, psychologists, social workers, etc), scientists (neuroscientists, cognitive psychologists, etc), and directors of schools specializing in learning and developmental disorders, as well as technology consultants. We integrated their feedback at different stages throughout the development of the platform. Regarding user experience, both structured and informal user feedback collection has been iterative and ongoing throughout the development process. The MindLogger apps described in this paper use an assessment frequency, duration, and question content that has been substantially validated over recent years by members of our team for a wide range of age groups (children, adults, and older adults) [62-66], for healthy individuals [62,63,67], and for persons with diverse forms of mental [62,63,67-71] or physical [72-74] disorder. The NIMH's predecessor to the MindLogger applet described below (using a highly similar protocol for daily assessments and with identical question content) documented an average completion rate of 77.9% (SE 0.81%) for repeated daily assessments, with no significant fatigue effect, defined as an increase in missing data as a function of time in the study (ordinary least squares linear model β coefficient=–.041; P<.001). In addition to user experience considerations and other stakeholder feedback, the information we gathered over the last 3 years that preceded the above review of customizable, mobile, experience sampling products also helped to guide MindLogger development.

Application of MindLogger in the Healthy Brain Network Study

To obtain more in-depth information on real-time tracking of emotions, behavior, daily activities, and their contextual influences in the Healthy Brain Network study [54], we adapted the combined actigraphy and experience sampling mobile assessment tools and content from the NIMH Family Study of Affective Spectrum Disorders [55]. Although the findings of that study (summarized in the Introduction section) were primarily based on adult samples, the inclusion of a substantial subset of offspring aged 10-18 years of parents with mood disorders and controls provided compelling evidence for the feasibility, acceptability, and clinical significance of experience sampling in youth. Therefore, the goal of the present initiative was to create a version in MindLogger, with updated content (particularly with regard to sleep, positive and negative thoughts, food and drink, internet, and social media), enhanced with clarification of the content, inclusion of colorful images, and formats adapted for children and young adults to encourage engagement [75].

Results

Overview

In this section, we present the results of our review of customizable, mobile, experience sampling products, the current state of the MindLogger platform (including roles and permissions, software architecture, and current set of features), and the applet we have deployed in the Healthy Brain Network study.

Results of the Review of Customizable, Mobile, and Experience Sampling Products

Our search resulted in 392 products, of which 315 appear to be in current use. Multimedia Appendix 1 contains a list of 101 products that have Android and/or iOS mobile apps. Upon closer inspection of their websites, 59 appeared relevant to scheduling questionnaires and notifications for a group of respondents, so we contacted the 59 products' companies/organizations through their web-based contact forms or via email to clarify their products' capabilities and exchanged emails with the 47 companies that responded. On the basis of these exchanges, we were able to identify 21 products that appeared to satisfy our primary criteria (administrator interface for creating and scheduling recurring, customized questionnaires, where users receive and respond to scheduled notifications on a mobile device). Five potential candidates were omitted as they are intended for use by internal business employees of a company, not by patients or by participants of a study, and require individual licenses, log-ins, or fees per device. Two more candidates were omitted as they required a legal agreement to demonstrate their products. Of the 14 remaining products for which we engaged in demonstrations and free trials, 10 met the primary criteria and are included in Figures 1-3. In these figures, *yes* or *no* (filled vs empty) correspond to the questions in the bulleted lists below. The questions represent specific instances of broader criteria that can be complex and nuanced.

- *Multiuser*: Can more than one end user access the app on the same mobile device, even if it means logging out and logging back in again?
- *Offline notification*: Can end users receive and respond to notifications without an internet connection?
- *Tap notification*: When an end user taps on a notification in their mobile device's notification bar, does it take them directly to their scheduled activity within the app?
- *Visualize data*: Is there a data visualization dashboard to review any individual end user's response data?
- *Reviewers*: If there is a data visualization dashboard, can an administrator give someone access to review only one end user's response data in the dashboard?
- *Setup languages*: When administrators create a customized questionnaire, can they choose from at least 5 different languages to use the interface, in addition to English? (This is distinct from how many different languages the end users can see.)
- *API*: Is there a consumer-facing application programming interface (API)?
- *Open*: Is the product's mobile app software fully open source and free to set up?

Figure 1 contains information about access to the product, software, and data (for both Android and iOS).

Figure 1. Access to experience sampling products, software, and data (filled means yes). API: application programmer interface.

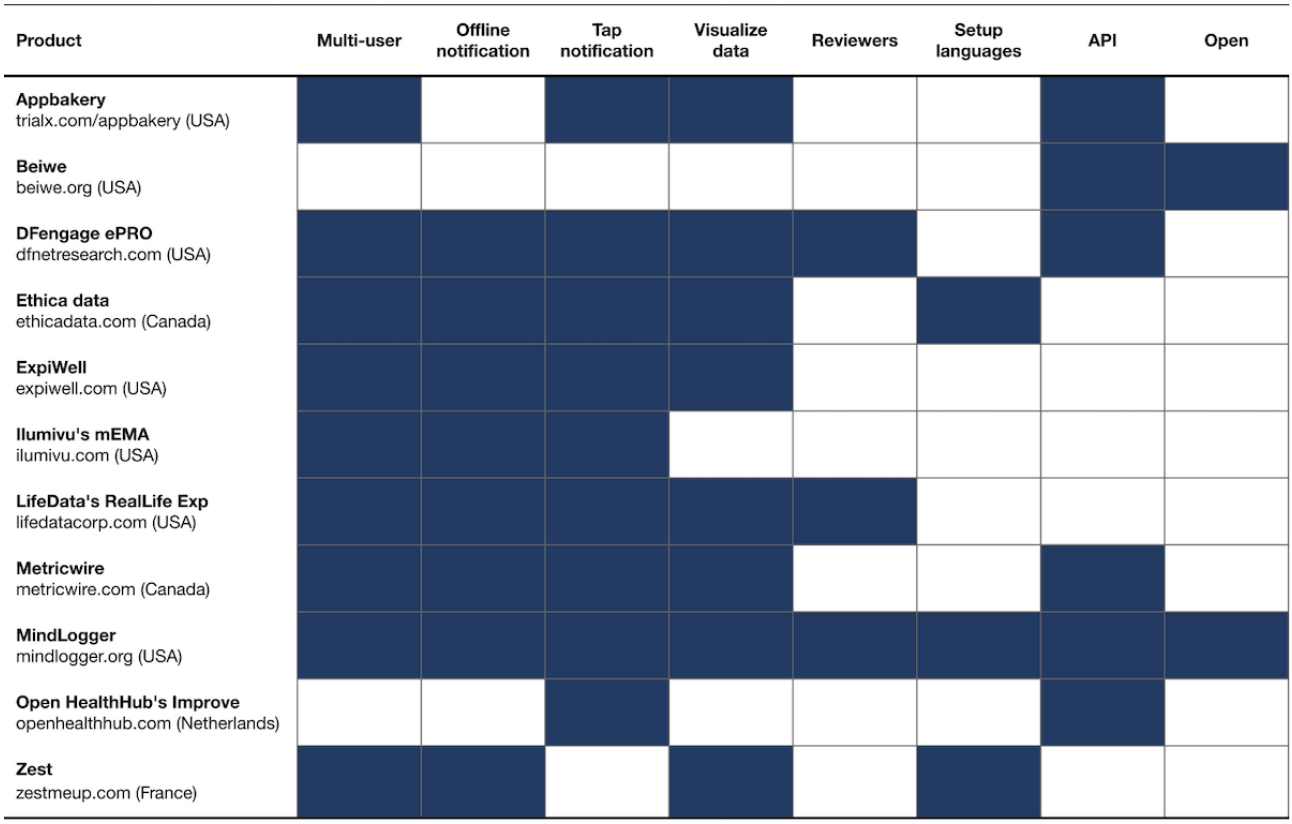


Figure 2 contains information about presentation and data capture features:

- *Play video/audio*: When administrators create an activity, can they include video or audio clips?
- *Type text, take photo, record video/audio, draw*: When administrators create an activity, can they include any of the following to capture data from end users? (Text entry, camera photo, audio/video recording, and drawing.)
- *GPS*: Can GPS location data be acquired through the app?
- *Sensor*: Can any additional sensor (eg, accelerometer) data be acquired?
- *Time*: Can any question include a countdown or timer?
- *Logic*: Can the response to a question determine which is the follow-up question (skip/branch logic)?
- *Score*: Can a questionnaire's scoring logic be entered when creating the questionnaire?

Figure 2. Presentation and data capture features of experience sampling products (filled means yes).

Product	Play video	Play audio	Type text	Take photo	Record video	Record audio	Draw	GPS	Sensor	Time	Skip	Score
Appbakery	■	■	■	■	■	■		■	■		■	
Beiwe			■			■		■	■		■	
DFengage ePRO			■								■	
Ethica data	■		■	■	■	■		■	■		■	
ExpiWell	■	■	■	■	■	■		■	■		■	
Ilumivu's mEMA	■	■	■	■	■	■		■	■	■	■	
LifeData's RealLife Exp	■	■	■	■				■			■	
Metricwire	■	■	■	■	■	■		■	■	■	■	■
MindLogger	■	■	■	■	■	■	■	■		■	■	■
Open HealthHub's Improve	■	■	■	■					■		■	■
Zest			■									

Figure 3 contains information about privacy and security:

- *Encrypt on (device, server)*: Does the product encrypt data on the mobile device or on the server?
- *End-to-end encryption*: Is the product end-to-end encrypted or could someone within the company or organization hosting apps on their server see (even anonymized respondents') response data?
- *Admin delete*: Can an administrator delete an individual end user's data without having to make a request from the product creator?
- *Own server*: Can the product be hosted on an administrator's own server?

Figure 3. Privacy and security of experience sampling products (filled means yes).

Product	Encrypt on device	Encrypt on server	End-to-end encryption	Admin delete	Own server
Appbakery					
Beiwe	■	■			■
DFengage ePRO	■	■	■	■	■
Ethica data	■	■		■	■
ExpiWell		■			
Ilumivu's mEMA	■	■		■	■
LifeData's RealLife Exp	■	■			
Metricwire	■	■	■	■	■
MindLogger	■	■	■	■	■
Open HealthHub's Improve	■	■	■		
Zest				■	

Current State of the MindLogger Platform

Roles and Permissions

Figure 4 shows a schematic of MindLogger, where an administrator selects, edits, or creates an applet; administers the applet to end users; and views or makes use of their data. Figure 5 outlines the different roles (owner, manager, coordinator, editor, user, or reviewer) and their permissions with regard to administration, content, use, and data management. For instance, a principal investigator of a research study could be the owner of a new applet as well as its editor and might assign a laboratory manager who recruits participants (applet users) to be the manager of the applet, and 2 data analysts to be the reviewers of all deidentified participants' data. As a second example, a clinical director of a pediatric mental health clinic could be the owner of a copy of an applet containing assessments, assigns a clinical coordinator who manages patients (applet users) to be a manager of the applet, and assigns parents to be reviewers of their child's responses. As a trivial example, anyone could sign up for a MindLogger account and own their own applet and be its only user. This last scenario enables any user to customize their own applet to send scheduled notifications to themselves to remind themselves to take their medication, perform breathing exercises, practice mindfulness meditation, or log their thoughts within the app.

Figure 4. MindLogger schematic. Left: select, edit, or create activities with scheduled notifications, such as questionnaires, tasks, or interventions, for use on mobile (iOS, Android) devices or the web. Middle: assign yourself or others to do these activities for data collection, annotation, research, or remote clinical assessment or therapy. Right: view end user data for which you have access.

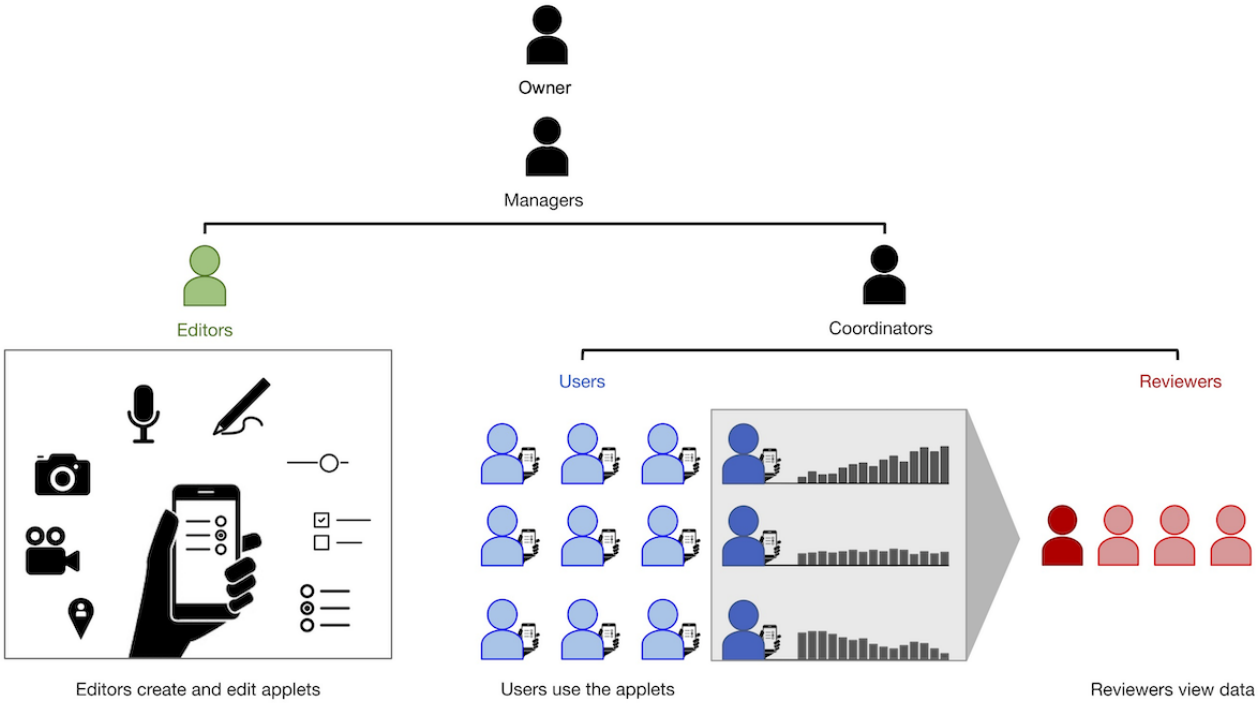
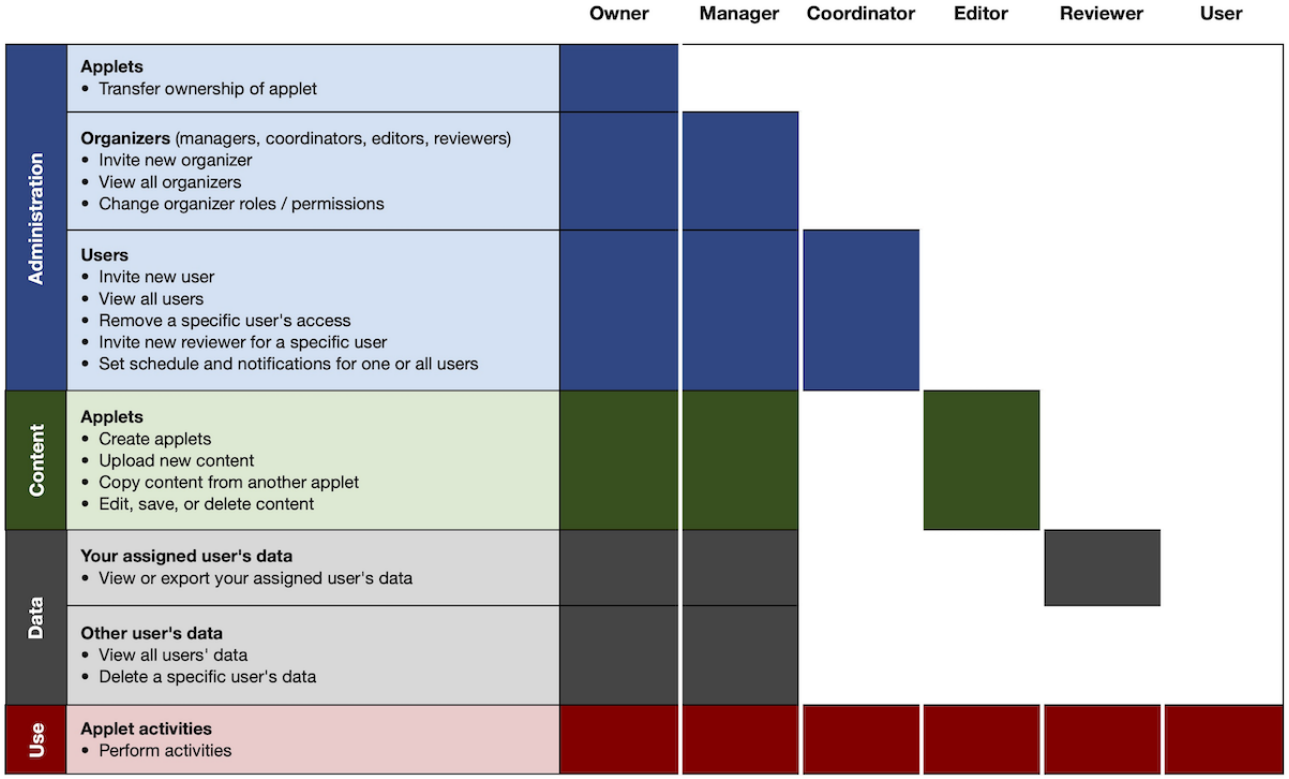


Figure 5. MindLogger roles and permissions for administration, content, use, and data management.



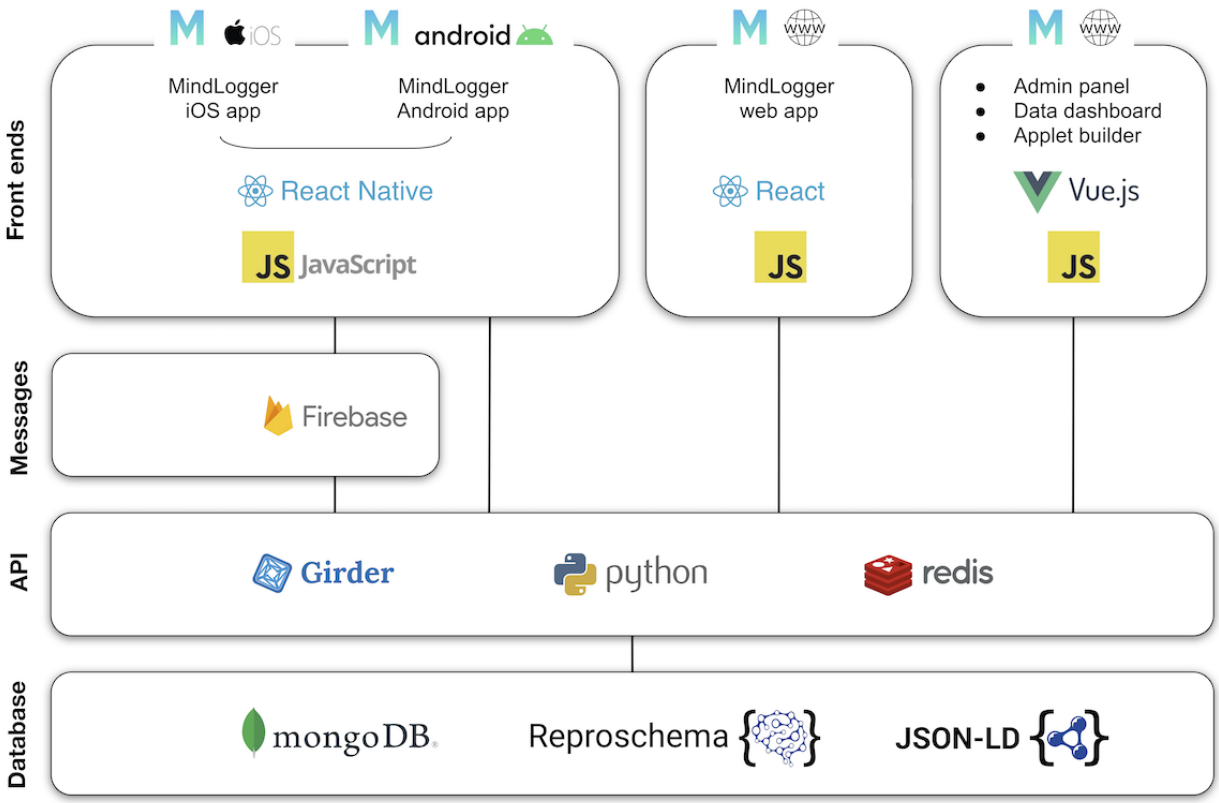
Administering permissions

1. All roles and associated permissions are restricted to a specific applet.
2. A person can have multiple roles for an applet, and different roles for different applets.
3. A person can become a User of an applet only via an invitation.
4. A person can become a Manager, Coordinator, Reviewer, or Editor of an applet via an administrator website.
5. Coordinators cannot invite themselves to be a Manager, Reviewer, or Editor unless they are the applet's Owner.

MindLogger Software Architecture

MindLogger's software architecture (Figure 6) consists of a set of end user–facing front ends (2 mobile apps and a web application) and organizer-facing front ends (an admin panel, data dashboard, and applet builder) with a shared RESTful HTTP API using MongoDB [76] for data storage. The mobile front ends, Android, and iOS apps are built using React Native [77]. This allows us to share a single code base across mobile platforms, resulting in increased speed of development and ease, and cost-effectiveness of maintenance. The web application is a ReactJS browser-based counterpart to mobile apps and currently provides a subset of their functionality. Administrators (managers, coordinators, editors, and reviewers) have access to different single-page applications built using VueJS [78]. The admin panel and applet builder manage user roles and applets, and the data dashboard is used to review user data, with custom charts implemented using d3.js [79]. The computer security firm Alpine Security [80] conducted extensive cybersecurity black, gray, and white penetration tests to ensure that MindLogger follows best practices for privacy and security. These practices can be adapted to the specific regulations and guidelines of different countries, including the General Data Protection Regulations of the European Union [81], which are among the strictest concerning data use, access, and storage. As shown in Figure 3, MindLogger has end-to-end encryption, permits administrators to delete an individual's data, and can be set up on one's own server (accommodating European regulations concerning the physical location of data processing and storage [81]), so the platform should already meet the security requirements of most use cases.

Figure 6. MindLogger software architecture diagram. API: application programmer interface.



Our application front-end and backend code base is accessible as web-based GitHub repositories [59] and is licensed under an extremely permissive Open Source Initiative–approved open-source license, the Common Public Attribution License (CPAL-1.0) [82]. The license requires that attribution be given by including (1) the copyright notice: "Copyright (c) 2017 MATTER Lab at the Child Mind Institute," (2) the MATTER Lab website address, (3) the Child Mind Institute's logo, and (4) the attribution phrase: "Child Mind Institute product intended for building applications for good." We include the attribution phrase to give credit to the developers while also making it clear that although we intend for people to build MindLogger applets that will be benevolent, we have no control over their intent, content, or the data they collect [83].

Our administrative software is licensed under a new license, the Delayed Open Source Attribution License (DOSA-1.0) [84]. Although it is not itself an open-source license, the purpose of this Delayed Open Source Attribution License is to provide open access to software for noncommercial use while giving attribution to its original developer, and after a delay of 3 years, forcing the software to fall under the terms of the open source CPAL-1.0 license that preserves the attribution information of this license. This delay is intended to protect the commercial interests of the licensor without compromising on the many benefits of creating open-source products [85].

The backend API is built in Python using Girder's RESTful API [86] with the CherryPy framework [87]. This software layer provides a set of RESTful endpoints that allow users, applets, activities, items (such as individual questions), and user responses. All user data are stored in a MongoDB database hosted in an Amazon Web Services [88] cloud instance with password-based encryption. Specifically, user response data are encrypted using their own password on the client side so that only managers or reviewers can view their data using an applet password; other sensitive information (name and email) is encrypted on the server side. We have Health Insurance Portability and Accountability Act compliance agreements with Amazon Web Services, Google Cloud Platform [89], and MongoDB Atlas [90], and the software permits installation on an arbitrary backend server (eg, on a university or hospital server and not on any cloud service provider's servers). For improved performance, MindLogger uses a Redis [91] instance as a temporary storage for data caching. MindLogger uses Firebase Cloud Messaging to send notifications from the backend server to the end user's mobile device. All additional data are consumed from the backend API through the HTTPS requests.

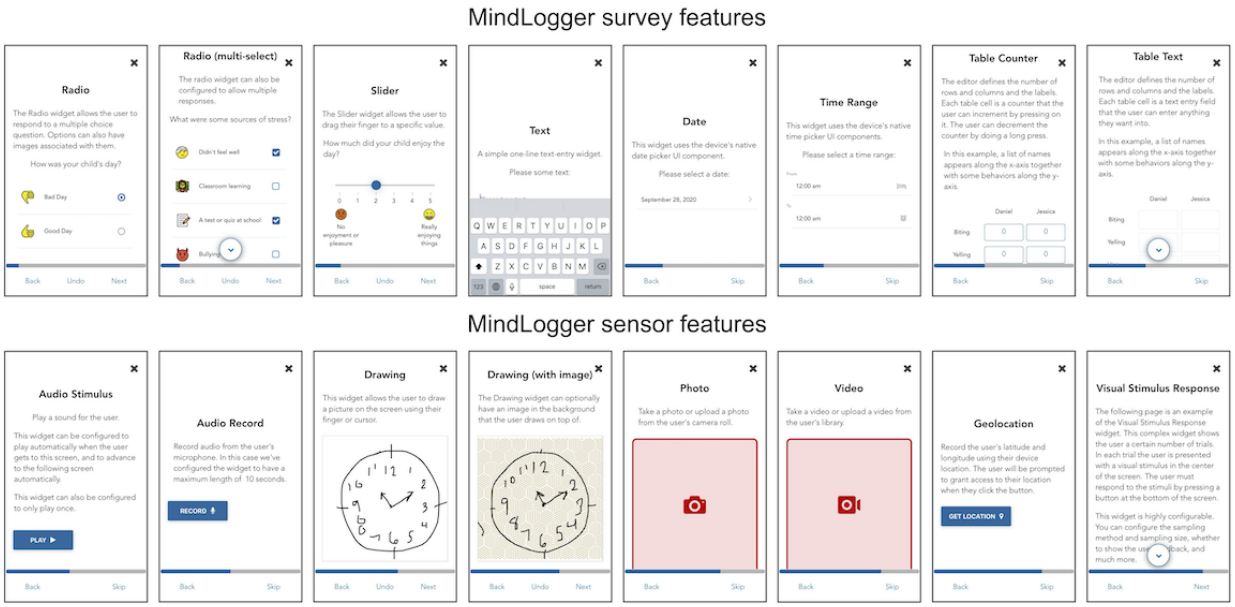
The applets, activities, and items are described using ReproSchema [92], an emerging standard for capturing and harmonizing cognitive, clinical, and behavioral assessments and responses in a provenance-preserving manner. The schema uses JSON for linking data (JSON-LD) [93] as its representation format and captures, as a connected graph of information, the details of the questions, presentation logic on the basis of responses or scheduling, computation of scores, and interface hints for applications such as MindLogger. The schema uses GitHub to maintain versions and provide persistent Uniform Resource Identifiers for applets and activities, supports multilingual applets, and uses World Wide Web Consortium

provenance specifications [94] to establish provenance between the response, the responder, and the applet.

MindLogger Current Features

We have succeeded in implementing many user- and administrator-facing features (see the MindLogger website [59] for updated information about features, installation, administration, and use). There are an arbitrary number of activities in an applet and an arbitrary number of screens in an activity (with response-based conditional logic directing the sequence of screens). Each screen of a MindLogger applet can display text and a picture or video, play a sound file, and present an interactive component with different possible response options, such as single- or multiple-selection check boxes, image selection, slider bar, text entry, table text or number entry, audio recording, photo or video capture, drawing or tapping on images, or GPS location button. Response delay and timer options are also available for each screen. MindLogger is cross-platform (iOS, Android, and web browser compatibility) and has open-source code for apps and applets, and data are end-to-end encrypted. The browser-based administration panel enables user management, easy creation, and customization of one's own mobile or web applications without programming or design experience, scheduling of applets and notifications per activity per user or group of users, and visualization and export of data. Figure 7 shows screenshots of the MindLogger mobile phone app features. We have created applets to remotely administer assessments as well as therapies and are currently constructing a public library with over 100 mental health and cognitive assessments that have open licenses for general use.

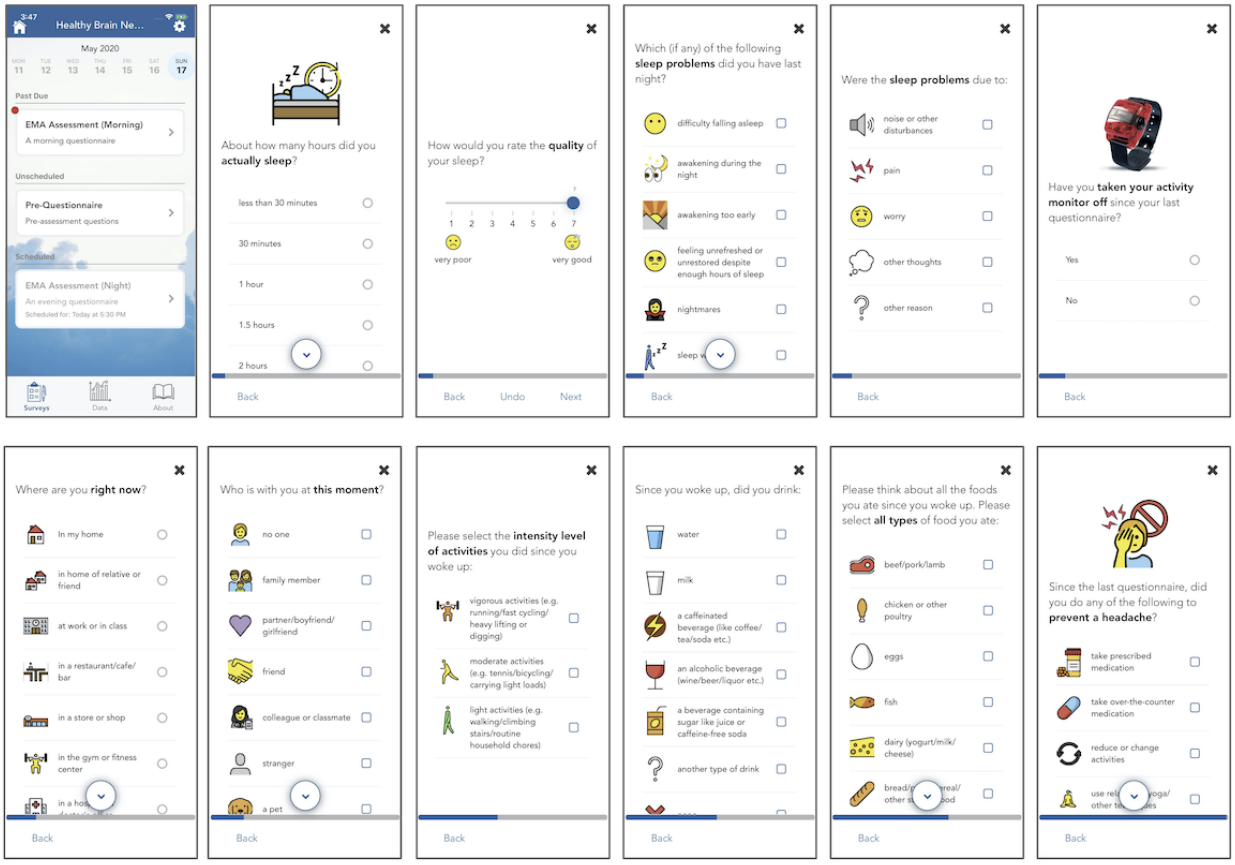
Figure 7. MindLogger screenshots showing survey and sensor features.



Deployment of MindLogger in the Healthy Brain Network Study

MindLogger's *NIMH-EMA* applet has been launched as part of the Healthy Brain Network study [54] to a vulnerable transdiagnostic New York City community sample (current n=4315; enrollment rate: 90 per month; >90% have mental health or learning disorders). Participants of the study receive multiple notifications per day on their Android or iOS devices to respond to morning, afternoon, and evening assessments. Figure 8 shows screenshots of the NIMH-EMA applet. We are currently enrolling children and adolescents who are at least 11 years old to use the NIMH-EMA applet as part of the Healthy Brain Network study. This applet is about to be deployed in the NIMH research program as well.

Figure 8. MindLogger National Institute of Mental Health-ecological momentary assessment applet screenshots.



EDUCATE Study's Daily and Weekly Diary Applet

Having developed and deployed the NIMH-EMA applet as part of one research study, we were able to rapidly develop and deploy a second assessment applet as part of a different research study on reading disability. Reading disability is the most common learning disability, affecting 10%-15% of school-age children [95]. It causes major functional impairments at all stages of life. A wealth of data documents lifelong disadvantages in educational and occupational attainment. Current evidence-based reading interventions largely rely on services by trained specialists, either in well-resourced classrooms or clinical settings. As such, under-resourced schools (or regions) are often unable to provide reading interventions for their students. The significance of this dilemma is compounded when considering that children of lower socioeconomic status and children with other serious comorbid behavioral health conditions may have more severe or complex reading disability profiles [96-98]. Thus, the children most in need are the least likely to have access to evidence-based treatment.

The EDUCATE study, funded by the National Institute of Child Health and Human Development, is a collaborative effort between researchers at the University of Connecticut and the Child Mind Institute. This clinical trial will examine the effectiveness of an at-home, game-based intervention for reading disorders. Parents of participants in this study are completing daily and weekly assessments in MindLogger, which will allow researchers to assess the home environment and compliance with the intervention protocol throughout the study period. These data are critical for evaluating the impact of this clinical trial.

Discussion

Principal Findings

In this paper, we reviewed customizable, mobile experience sampling products for configurable data collection and content delivery, summarized the motivation for and development of a new mobile platform called MindLogger, and described an initial use case that applies a MindLogger version of the NIMH Family Study app as part of the Healthy Brain Network study.

Our review returned an initial set of 392 products, of which 59 appeared to be in current use, had Android and iOS mobile apps, and were capable of scheduling questionnaires and notifications for a group of respondents. Of the 47 companies that responded to our inquiries and did not require a legal agreement, we identified 10 products that satisfied our primary criteria (administrator interface for creating and scheduling recurring, customized questionnaires, where an arbitrary number of users can receive and respond to scheduled notifications on their mobile devices). Four of these products supported end-to-end encryption, 2 enabled restricted access to individual users' data, 1 provided open-source software, and none provided all 3 of these capabilities. Our review is not exhaustive regarding either the products that could possibly be used for experience sampling or the range of features that these products support. Existing

products include an impressive assortment of nonoverlapping features and complement each other in ways that reflect the different niches or markets for which they were intended. The limitations of our search for existing products are as follows: (1) our search queries and Google's search algorithm may not reflect the optimal search criteria for some relevant products; (2) product websites are sometimes unclear and incomplete and may even misrepresent the capabilities of their products; and (3) some companies did not respond to questions even through their website's question or support page. Products are also adding features over time, and some companies or organizations offer paid services to build the desired features. Our comparison should be seen as a snapshot of the current state of a subset of features provided by software products intended for experience sampling.

We successfully developed MindLogger, a new platform that meets most of the stated needs of our collaborators around the world, who desire an open source, mobile mental health platform to inform, assess health, acquire data, and administer therapies. We prioritized clinicians' and researchers' needs and users' experience during development, aligned technologies to meet these priorities, and are on track to achieve the full feature set we set out to include. MindLogger has end-to-end encryption, enables restricted access, is open source, and supports a variety of data collection features. One limitation of MindLogger is that it currently does not support passive data collection or interaction with peripheral devices. This was intended to reduce concerns about surveillance in an app whose first use case was for assessing children and adolescents. However, in the future, we intend to support passive monitoring of location and behaviors, and communication with other devices, so long as these are opt-in by the end user and there are clear reminders to the end user concerning what data are collected and how they will be used. Another potential limitation lies in its core strength: by making MindLogger flexible, modular, cross-platform, and configurable to help meet the unforeseen needs of future applet builders and users, the creation of variants of even well-vetted instruments is likely and will necessitate their careful validation. We have recently built a web-based *applet library* for viewing, copying, editing, and sharing applets. Applets in the library are labeled to indicate which have been created or vetted by the Child Mind Institute and which have been contributed by others.

We have demonstrated the flexibility and applicability of the MindLogger platform through the deployment of the NIMH-EMA applet in the Healthy Brain Network, a large-scale, longitudinal, mobile mental health study. The NIMH is about to launch the same applet in its own research program. For future directions in the near term, there are a variety of other mental health–related applets in preparation for deployment and in the planning stages as described below.

MindLogger Applets in Preparation for Deployment

We are currently refining and testing MindLogger applets to assess and administer interventions targeting specific subgroups of youth with particular mental health and learning disorders. Although some of these applets support specific collaborators' research, others are for broader use (listed below, with video screencast demonstrations on the MindLogger website [59]).

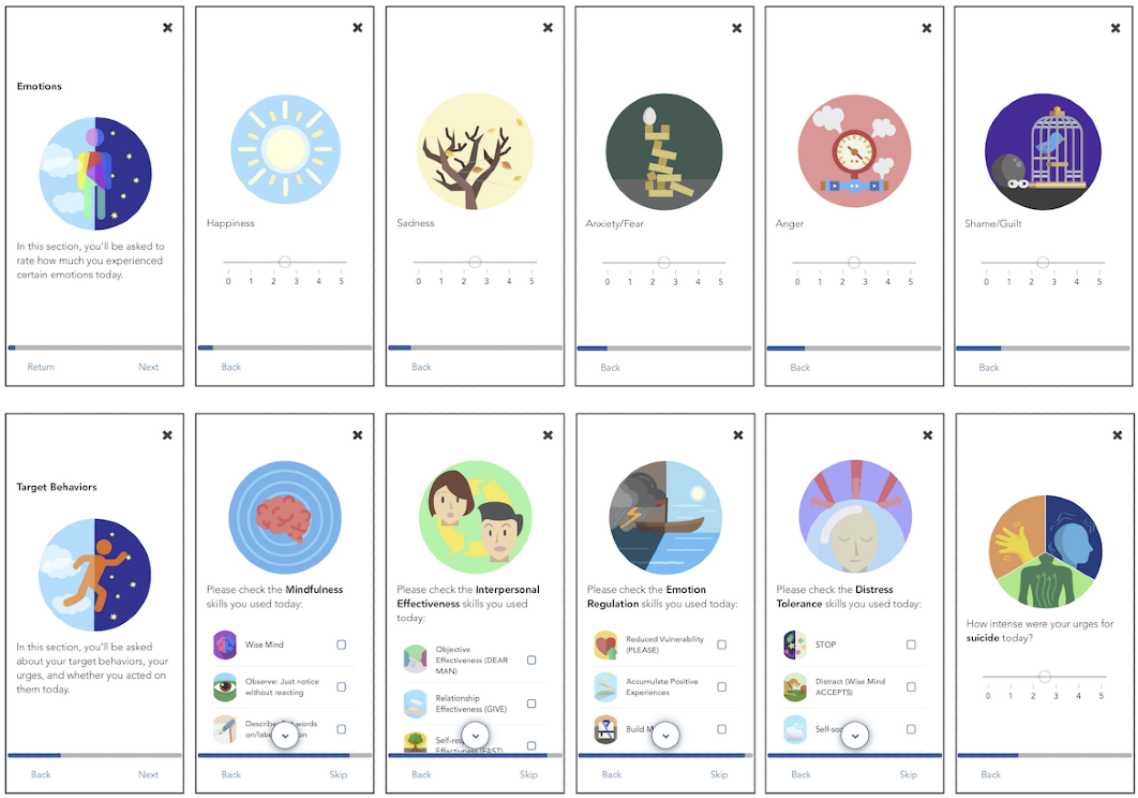
Pediatric Screener Mental Health Screening Applet

Integrating primary care and mental health has been associated with improved patient outcomes [99]; therefore, mental health screening in pediatric clinics could lead to earlier diagnosis and improved outcomes for patients. The Hearst Foundations supported the development of a pediatric screener tool using MindLogger. This tool will administer assessments to children or their parents, for children receiving a wellness checkup at their pediatric clinic, and alert their physician if a child shows signs of a mental health disorder. We have built the applet and will pilot it at the Richmond University Medical Center in Staten Island, New York. The initial screening questionnaire assesses internalizing and externalizing symptoms, issues of attention and hyperactivity, depression and suicidal ideation, disordered eating behavior, and experiences of bullying. It also collects demographic information about the child and parent. Children with a clinically significant level of symptoms were prompted to complete additional questionnaires to collect more detailed information. A similar questionnaire was piloted in several New York City–based pediatric settings and found it to be an effective tool for identifying children at risk of a serious mental health disorder. The MindLogger platform will create a much more streamlined process and user-friendly experience, increasing the probability of adoption by more pediatric practices and clinics.

Dialectical Behavior Therapy Applet

The Dialectical Behavior Therapy (DBT) Diary Card is a digitized version of the diary card used in evidence-based DBT programs. This tool is a daily tracker of mood, targeting behavioral urges and specific behaviors, and the use of coping skills to manage these emotions, urges, and behaviors. Our DBT applet (Figure 9) will enable therapists to create a digital DBT diary card with a patient to include specific treatment targets for that individual. This will allow patients to set notification reminders to complete the diary card on a daily basis. The patient's device will automatically update with the new targets and schedule, allowing the user to progress in the therapy without having to change the way the data are retrieved. Both patients and therapists will have access to the diary card data to guide treatment planning and sessions.

Figure 9. MindLogger Dialectical Behavior Therapy applet screenshots.



TokenLogger Behavior Intervention Applet

We are refining a behavior intervention applet called *TokenLogger* that can be customized to retrospectively and prospectively track specific behaviors and help promote and reinforce desirable behaviors while reducing or extinguishing undesirable behaviors. It will also help track the frequency, duration, and severity of target behaviors to inform modifications in behavioral treatment plans and assess progress toward outcomes and goals. The TokenLogger applet will be evaluated by clinical experts at the Child Mind Institute and piloted with patients at the Child Mind Medical Practice, to better understand the timing, duration, and frequency of undesirable behaviors, and to test the efficacy of this digital rendition of behavior modification therapy.

Planned MindLogger Applets

We intend to replace all of the Child Mind Institute's and the Child Mind Medical Practice's pencil-and-paper assessments with MindLogger applets and, where appropriate, create and test digital renditions of therapies as MindLogger applets, and make these universally accessible as part of a web-based public library of MindLogger applets for anyone to use, modify, and translate. In addition, there are a few applets in the planning stages.

Diagnostic Screening Applet for Kiddie Schedule for Affective Disorders and Schizophrenia

We are currently developing a MindLogger applet with part of the Composite International Diagnostic Interview Screener that has been incorporated into the NIMH version of the epidemiologic version of the Kiddie Schedule for Affective Disorders and Schizophrenia. This screener is being tested for parents and self-administration in order to streamline the process as we transition to more automated approaches for large-scale studies of youth.

Taction Exposure Therapy Applet

We created a prototype iOS and Android mobile app called *Taction* that is a simple exposure therapy game for children who have obsessive-compulsive disorder or anxiety-related issues. The app rewards users for tapping on images that heighten their anxiety, potentially helping them progress in their treatment between exposure therapy sessions. Once incorporated into MindLogger, the Taction applet will be evaluated at the Child Mind Institute and piloted with patients at the Child Mind Medical Practice to assess the relative efficacy of a digital rendition of exposure therapy.

NIMH Cognitive Task Battery

In the same manner as we want to enable anyone to create, configure, and administer their own mobile questionnaires, we also want to enable anyone to do the same for different types of cognitive assessments. In this spirit, we created a Flanker task applet that is thoroughly configurable (presentation and timing of a fixation target, stimulus, and feedback), and we are collaborating with the NIMH to create a small battery of cognitive tasks for research in modeling behaviors.

Acknowledgments

This project was supported by the Hearst Foundations, the Intramural Research Program of the National Institute of Mental Health (grants with principal investigator KRM: ZIA MH002954 Motor Activity Research Consortium for Health and ZIA MH002804-15 Family Study of Affective and Anxiety Spectrum Disorders); the Centre for Research and Interdisciplinarity, Université de Paris; the Blanca and Sunil Hirani Foundation, Massachusetts Institute of Technology through the laboratory of SSG (SSG and SAA were supported by grants: NIH P41 EB019936) and Joseph Healey. The authors thank the companies and organizations that responded to the questions and provided demonstrations of their products. The authors thank Rangle.io for their help in refactoring MindLogger's mobile app code base and Alpine Security for conducting penetration tests of the entire code base. They also thank Paul Mitrani and Stephanie Samar for clinical feedback during the development of MindLogger applets and volunteers Chynna Levin, Mairav Linzer, and Benjamin Vogel. Finally, the authors would like to thank all of their colleagues and collaborators around the world who have encouraged and helped guide the development of MindLogger in anticipation of using it for their many different data collection, research, clinical, and other endeavors.

A Krishnakumar was affiliated with the MATTER (Mind-Assisting Technologies for Therapy, Education, and Research) Lab at the Child Mind Institute and with the Centre for Research and Interdisciplinarity, Université de Paris at the time of this study and is currently affiliated with the ETH Library Lab, ETH Zurich, and Citizen Science Centre in Zurich. JC and A Keshavan were affiliated with the MATTER Lab at the Child Mind Institute for part of this study; JC is currently affiliated with the Computational Neuroimaging Laboratory at the Child Mind Institute, and A Keshavan is currently affiliated with Octave Bioscience.

Authors' Contributions

A Klein founded and led the MindLogger project, led this study, wrote the manuscript, and contacted all of the companies or organizations as part of the product comparison that led to Figures 1-3. A Klein, A Krishnakumar, AB, and KK collected information for Figures 1-3 (detailed in Multimedia Appendix 1). SSG and SAA lead the Repronim effort underlying MindLogger's data schema. A Krishnakumar helped to identify candidate features for MindLogger based on discussions with different stakeholders. JC, BT, IS, NA, A Keshavan, and HR contributed to the software engineering of MindLogger. WVA helped to direct all aspects of MindLogger software development and design as MindLogger's project manager. KRM and JS contributed to the MindLogger project through content developed by the NIMH Family Study of Affective Spectrum Disorders. YX helped to coordinate content in the National Institute of Mental Health-ecological momentary assessment (NIMH-EMA), Dialectical Behavior Therapy (DBT), and Kiddie Schedule for Affective Disorders and Schizophrenia applets, and found, modified, and designed images and icons for the NIHM_EMA, DBT, and EDUCATE applets. SS conducted repeated, extensive quality assurance of MindLogger functionality. AB and A Krishnakumar managed MATTER Lab volunteers in the United States and in France. LMA managed research collaborations as Research Operations Manager for the Child Mind Institute. ABL and A Krishnakumar oversaw formulation of MindLogger requirements relevant to Centre for Research and Interdisciplinarity (CRI) projects. MM and SSG provided consultation. KRM, JS, JC, LMA, WVA, BT, IS, NA, and MM reviewed and modified the manuscript for scientific content. AB helped in formatting the manuscript.

Conflicts of Interest

While currently employed by Octave Bioscience, A Keshavan was not an employee while contributing to this manuscript.

Multimedia Appendix 1
Protocol for product search.
[DOCX File , 215 KB - jmir_v23i11e22369_app1.docx]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
2. Merikangas KR, He J, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. Pediatrics 2010 Jan;125(1):75-81 [FREE Full text] [doi: 10.1542/peds.2008-2598] [Medline: 20008426]
3. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization world mental health surveys. J Am Med Assoc 2004 Jun 02;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
4. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
5. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 08;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]

6. Raine R, Carter S, Sensky T, Black N. 'Referral into a void': opinions of general practitioners and others on single point of access to mental health care. J R Soc Med 2005 Apr;98(4):153-157 [FREE Full text] [doi: 10.1258/jrsm.98.4.153] [Medline: 15805555]
7. Ward-Ciesielski EF, Rizvi SL. Finding mental health providers in the United States: a national survey and implications for policy and practice. J Ment Health 2019 Oct 24:1-7. [doi: 10.1080/09638237.2019.1677867] [Medline: 31647364]
8. Malowney M, Keltz S, Fischer D, Boyd JW. Availability of outpatient care from psychiatrists: a simulated-patient study in three U.S. cities. Psychiatr Serv 2015 Jan 01;66(1):94-96. [doi: 10.1176/appi.ps.201400051] [Medline: 25322445]
9. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Committee to Evaluate the Department of Veterans Affairs Mental Health Services. Evaluation of the Department of Veterans Affairs Mental Health Services. Washington, DC: National Academies Press; 2018.
10. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
11. Lefebvre A, Sommerauer J, Cohen N, Waldron S, Perry I. Where did all the "No-shows" go? Can J Psychiatry 1983 Aug;28(5):387-390. [doi: 10.1177/070674378302800512] [Medline: 6627199]
12. Campbell DG, Downs A, Meyer WJ, McKittrick MM, Simard NM, O'Brien P. A preliminary survey of pediatricians' experiences with and preferences for communication with mental health specialists. Fam Syst Health 2018 Sep;36(3):404-409. [doi: 10.1037/fsh0000309] [Medline: 29199842]
13. Sipe M. The effects of stigma toward mental illness on family physicians. Thesis and Dissertations - The University of Arizona, Tucson, Arizona. 2016. URL: https://repository.arizona.edu/handle/10150/603680 [accessed 2021-10-07]
14. Ginn SK, Clark E. The medical profession and stigma against people who use drugs. Br J Psychiatry 2017 Dec;211(6):400. [doi: 10.1192/bjp.211.6.400] [Medline: 29196405]
15. Klein JD, McNulty M, Flatau CN. Adolescents' access to care: teenagers' self-reported use of services and perceived access to confidential care. Arch Pediatr Adolesc Med 1998 Jul;152(7):676-682. [doi: 10.1001/archpedi.152.7.676] [Medline: 9667540]
16. Kuramoto-Crawford SJ, Han B, McKeon RT. Self-reported reasons for not receiving mental health treatment in adults with serious suicidal thoughts. J Clin Psychiatry 2017 Jun;78(6):631-637. [doi: 10.4088/JCP.16m10989] [Medline: 28406268]
17. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
18. Dunster GP, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. Neuropsychopharmacology 2021 Jan 12;46(1):197-208. [doi: 10.1038/s41386-020-00830-5] [Medline: 32919408]
19. Demaerschalk BM, Blegen RN, Ommen SR. Scalability of telemedicine services in a large integrated multispecialty health care system during COVID-19. Telemed J E Health 2021 Jan;27(1):96-98. [doi: 10.1089/tmj.2020.0290] [Medline: 32795147]
20. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
21. Loades ME, Chatburn E, Higson-Sweeney N, Reynolds S, Shafran R, Brigden A, et al. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. J Am Acad Child Adolesc Psychiatry 2020 Nov;59(11):1218-1239 [FREE Full text] [doi: 10.1016/j.jaac.2020.05.009] [Medline: 32504808]
22. Digital 2021 Global Overview Report (2021). DataReportal. URL: https://datareportal.com/global-digital-overview [accessed 2021-01-28]
23. 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]
24. Schueller SM, Begale M, Penedo FJ, Mohr DC. Purple: a modular system for developing and deploying behavioral intervention technologies. J Med Internet Res 2014 Jul 30;16(7):e181 [FREE Full text] [doi: 10.2196/jmir.3376] [Medline: 25079298]
25. Purple Robot. Center for Behavioral Intervention Technologies. URL: https://cbits.northwestern.edu/ [accessed 2020-07-07]
26. 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 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
27. The IntelliCare Platform. Center for Behavioral Intervention Technologies. URL: https://intellicare.cbits.northwestern.edu [accessed 2020-07-07]
28. Beiwe Research Platform. URL: https://www.beiwe.org/ [accessed 2020-07-07]
29. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
30. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using Researchkit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]

31. mPower 2.0. Parkinson mPower. URL: https://parkinsonmpower.org/your-story [accessed 2020-07-07]
32. ResearchKit and CareKit. Apple. URL: http://www.apple.com/researchkit/ [accessed 2020-07-07]
33. ResearchStack. URL: http://researchstack.org/ [accessed 2020-07-07]
34. M-health Index and Navigation Database (MIND). Beth Israel Deaconess Medical Center. URL: https://apps.digitalpsych.org/ [accessed 2020-07-07]
35. Torous JB, Chan SR, Yellowlees PM, Boland R. To use or not? Evaluating aspects of smartphone apps and mobile technology for clinical care in psychiatry. J Clin Psychiatry 2016 Jun;77(6):734-738. [doi: 10.4088/JCP.15com10619] [Medline: 27136691]
36. The App Evaluation Model. Psychiatry. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2020-07-07]
37. Larson R, Csikszentmihalyi M. The experience sampling method. In: Reis H, editor. Naturalistic Approaches to Studying Social Interaction. San Francisco, California, United States: Jossey-Bass Inc; 1983:41-56.
38. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
39. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. Clin Psychol Rev 2012 Aug;32(6):510-523. [doi: 10.1016/j.cpr.2012.05.007] [Medline: 22721999]
40. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020 Jan;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
41. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. Schizophr Bull 2008 May;34(3):507-514 [FREE Full text] [doi: 10.1093/schbul/sbm113] [Medline: 17932087]
42. Husky MM, Gindre C, Mazure CM, Brebant C, Nolen-Hoeksema S, Sanacora G, et al. Computerized ambulatory monitoring in mood disorders: feasibility, compliance, and reactivity. Psychiatry Res 2010 Jul 30;178(2):440-442. [doi: 10.1016/j.psychres.2010.04.045] [Medline: 20488558]
43. Johnson EI, Barrault M, Nadeau L, Swendsen J. Feasibility and validity of computerized ambulatory monitoring in drug-dependent women. Drug Alcohol Depend 2009 Jan 01;99(1-3):322-326. [doi: 10.1016/j.drugalcdep.2008.06.010] [Medline: 18692969]
44. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]
45. Serre F, Fatseas M, Debrabant R, Alexandre J, Auriacombe M, Swendsen J. Ecological momentary assessment in alcohol, tobacco, cannabis and opiate dependence: a comparison of feasibility and validity. Drug Alcohol Depend 2012 Nov 01;126(1-2):118-123. [doi: 10.1016/j.drugalcdep.2012.04.025] [Medline: 22647899]
46. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014 Dec 15;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]
47. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec;26(4):e1562 [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
48. Cognition measures. Health Measures. URL: https://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox/cognition [accessed 2020-07-06]
49. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):54-64 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872ded] [Medline: 23479546]
50. Technology. Neuroscape. URL: https://neuroscape.ucsf.edu/technology/ [accessed 2020-07-07]
51. Cambridge Cognition. URL: https://www.cambridgecognition.com [accessed 2020-07-07]
52. Ellison WD. An initial study of practicing psychologists' views of the utility of ecological momentary assessment for difficult psychotherapy cases. Adm Policy Ment Health 2021 Jul 12;48(4):597-607. [doi: 10.1007/s10488-020-01093-4] [Medline: 33047277]
53. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. Int J Bipolar Disord 2020 Dec 01;8(1):38 [FREE Full text] [doi: 10.1186/s40345-020-00201-5] [Medline: 33258015]
54. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data 2017 Dec 19;4:170181 [FREE Full text] [doi: 10.1038/sdata.2017.181] [Medline: 29257126]
55. Merikangas KR, Cui L, Heaton L, Nakamura E, Roca C, Ding J, et al. Independence of familial transmission of mania and depression: results of the NIMH family study of affective spectrum disorders. Mol Psychiatry 2014 Feb;19(2):214-219. [doi: 10.1038/mp.2013.116] [Medline: 24126930]

56. Lamers F, Swendsen J, Cui L, Husky M, Johns J, Zipunnikov V, et al. Mood reactivity and affective dynamics in mood and anxiety disorders. J Abnorm Psychol 2018 Oct;127(7):659-669. [doi: 10.1037/abn0000378] [Medline: 30335438]
57. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]
58. Johns JT, Di J, Merikangas K, Cui L, Swendsen J, Zipunnikov V. Fragmentation as a novel measure of stability in normalized trajectories of mood and attention measured by ecological momentary assessment. Psychol Assess 2019 Mar;31(3):329-339. [doi: 10.1037/pas0000661] [Medline: 30802118]
59. MindLogger. URL: https://mindlogger.org/ [accessed 2020-07-06]
60. Lyon AR, Munson SA, Renn BN, Atkins DC, Pullmann MD, Friedman E, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability. JMIR Res Protoc 2019 Oct 09;8(10):e14990 [FREE Full text] [doi: 10.2196/14990] [Medline: 31599736]
61. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
62. Bouvard A, Dupuy M, Schweitzer P, Revranche M, Fatseas M, Serre F, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. Am J Addict 2018 Oct;27(7):553-556. [doi: 10.1111/ajad.12804] [Medline: 30260085]
63. Dupuy M, Misdrahi D, N'Kaoua B, Tessier A, Bouvard A, Schweitzer P, et al. Mobile cognitive testing in patients with schizophrenia: a controlled study of feasibility and validity. J Thérapie Comportementale Cognitive 2018 Dec;28(4):204-213. [doi: 10.1016/j.jtcc.2018.02.002]
64. Rullier L, Atzeni T, Husky M, Bouisson J, Dartigues J, Swendsen J, et al. Daily life functioning of community-dwelling elderly couples: an investigation of the feasibility and validity of ecological momentary assessment. Int J Methods Psychiatr Res 2014 Jun;23(2):208-216 [FREE Full text] [doi: 10.1002/mpr.1425] [Medline: 24375556]
65. Salamon R, I.-Johnson E, Grondin O, Swendsen J. Étude des problèmes comportementaux et émotionnels chez l'adolescent : faisabilité et validité de l'approche esm. J Thérapie Comportementale Cognitive 2009 Mar;19(1):41-46. [doi: 10.1016/j.jtcc.2009.04.004]
66. Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep;26(3):e1521 [FREE Full text] [doi: 10.1002/mpr.1521] [Medline: 27539327]
67. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]
68. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. Schizophr Bull 2008 May;34(3):507-514 [FREE Full text] [doi: 10.1093/schbul/sbm113] [Medline: 17932087]
69. Husky MM, Gindre C, Mazure CM, Brebant C, Nolen-Hoeksema S, Sanacora G, et al. Computerized ambulatory monitoring in mood disorders: feasibility, compliance, and reactivity. Psychiatry Res 2010 Jul 30;178(2):440-442. [doi: 10.1016/j.psychres.2010.04.045] [Medline: 20488558]
70. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014 Dec 15;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]
71. Serre F, Fatseas M, Debrabant R, Alexandre J, Auriacombe M, Swendsen J. Ecological momentary assessment in alcohol, tobacco, cannabis and opiate dependence: a comparison of feasibility and validity. Drug Alcohol Depend 2012 Nov 01;126(1-2):118-123. [doi: 10.1016/j.drugalcdep.2012.04.025] [Medline: 22647899]
72. Jean FA, Sibon I, Husky M, Couffinhal T, Swendsen J. Feasibility and validity of ecological momentary assessment in patients with acute coronary syndrome. BMC Cardiovasc Disord 2020 Nov 27;20(1):499 [FREE Full text] [doi: 10.1186/s12872-020-01774-w] [Medline: 33246420]
73. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and acceptability of ecological momentary assessment of daily functioning among older adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840 [FREE Full text] [doi: 10.1016/j.jagp.2016.11.019] [Medline: 28065496]
74. Johnson EI, Sibon I, Renou P, Rouanet F, Allard M, Swendsen J. Feasibility and validity of computerized ambulatory monitoring in stroke patients. Neurology 2009 Nov 10;73(19):1579-1583. [doi: 10.1212/WNL.0b013e3181c0d466] [Medline: 19901250]
75. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
76. The application data platform. MongoDB. URL: https://www.mongodb.com [accessed 2020-07-06]
77. React Native · A framework for building native apps using ReactT. React Native. URL: https://reactnative.dev/ [accessed 2020-07-06]

78. The progressive JavaScript framework. VueJS. URL: https://vuejs.org/ [accessed 2020-10-01]
79. Data-driven documents. D3js. URL: https://d3js.org/ [accessed 2020-10-01]
80. Alpine Security. URL: https://alpinesecurity.com/ [accessed 2020-07-07]
81. Regulation (EU) 2016/679 of the European Parliament and of the Council. Official Journal of the European Union. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN [accessed 2021-01-28]
82. Common P. Common Public Attribution License Version 1.0 (CPAL-1.0). Open Source Initiative. URL: https://opensource.org/licenses/CPAL-1.0 [accessed 2020-07-07]
83. ChildMindInstitute / mindlogger-app. GitHub. URL: https://github.com/ChildMindInstitute/mindlogger-app/blob/master/LICENSE.md [accessed 2020-10-16]
84. Delayed Open Source Attribution License 1.0 (DOSA-1.0). URL: https://mindlogger.org/DOSA1 [accessed 2021-01-31]
85. ChildMindInstitute / mindlogger-admin. GitHub. URL: https://github.com/ChildMindInstitute/mindlogger-admin/blob/master/LICENSE.md [accessed 2021-01-31]
86. Girder: a data management platform. Girder: Read the Docs. URL: https://girder.readthedocs.io/en/stable/ [accessed 2020-07-06]
87. A minimalist Python web framework. CherryPy. URL: https://cherrypy.org/ [accessed 2020-10-01]
88. Amazon Web Services (AWS) - Cloud Computing Services. URL: https://aws.amazon.com/ [accessed 2020-07-06]
89. Cloud computing services. Google Cloud. URL: https://cloud.google.com/ [accessed 2020-07-06]
90. Managed MongoDB Hosting - Database-as-a-service. MongoDB. URL: https://www.mongodb.com/cloud/atlas [accessed 2020-07-06]
91. Redis. URL: https://redis.io/ [accessed 2020-10-01]
92. Reproschema Documentation. URL: https://www.repronim.org/reproschema/ [accessed 2020-10-01]
93. JSON for Linking Data. JSON-LD. URL: https://json-ld.org/ [accessed 2020-10-01]
94. Moreau L, Missier P. PROV-DM: The PROV Data Model. W3C. URL: https://www.w3.org/TR/2012/CR-prov-dm-20121211/diff.html [accessed 2020-10-01]
95. Moats LC, Dakin KE. Basic Facts About Dyslexia and Other Reading Problems. Baltimore, Maryland, United States: The International Dyslexia Association; 2007:1-134.
96. Duncan GJ, Magnuson K. Socioeconomic status and cognitive functioning: moving from correlation to causation. Wiley Interdiscip Rev Cogn Sci 2012 May;3(3):377-386. [doi: 10.1002/wcs.1176] [Medline: 26301469]
97. Francis DA, Caruana N, Hudson JL, McArthur GM. The association between poor reading and internalising problems: a systematic review and meta-analysis. Clin Psychol Rev 2019 Feb;67:45-60. [doi: 10.1016/j.cpr.2018.09.002] [Medline: 30528985]
98. Becker N, Vasconcelos M, Oliveira V, Santos FC, Bizarro L, Almeida RM, et al. Genetic and environmental risk factors for developmental dyslexia in children: systematic review of the last decade. Dev Neuropsychol 2017;42(7-8):423-445. [doi: 10.1080/87565641.2017.1374960] [Medline: 29068706]
99. Zeiss AM, Karlin BE. Integrating mental health and primary care services in the department of veterans affairs health care system. J Clin Psychol Med Settings 2008 Mar;15(1):73-78. [doi: 10.1007/s10880-008-9100-4] [Medline: 19104957]

Abbreviations

API: application programming interface
DBT: Dialectical Behavior Therapy
EMA: ecological momentary assessment
MATTER: Mind-Assisting Technologies for Therapy, Education, and Research
MRI: magnetic resonance imaging
NIMH: National Institute of Mental Health

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by H Vachon, F Bos, P Delespaule; comments to author 09.01.21; revised version received 03.02.21; accepted 16.03.21; published 11.11.21.

Please cite as:
Klein A, Clucas J, Krishnakumar A, Ghosh SS, Van Auken W, Thonet B, Sabram I, Acuna N, Keshavan A, Rossiter H, Xiao Y, Semenuta S, Badioli A, Konishcheva K, Abraham SA, Alexander LM, Merikangas KR, Swendsen J, Lindner AB, Milham MP
Remote Digital Psychiatry for Mobile Mental Health Assessment and Therapy: MindLogger Platform Development Study
J Med Internet Res 2021;23(11):e22369
URL: https://www.jmir.org/2021/11/e22369
doi:10.2196/22369
PMID:34762054

Original Paper

Practice Effects of Mobile Tests of Cognition, Dexterity, and Mobility on Patients With Multiple Sclerosis: Data Analysis of a Smartphone-Based Observational Study

Tim Woelfle[1,2], MSc, MD; Silvan Pless[1,2], MSc; Andrea Wiencierz[3], PhD; Ludwig Kappos[1,2], MD; Yvonne Naegelin[1,2], MD; Johannes Lorscheider[1,2], MD

[1]Research Center for Clinical Neuroimmunology and Neuroscience Basel, University of Basel, University Hospital Basel, Basel, Switzerland

[2]Department of Neurology, University Hospital Basel, Basel, Switzerland

[3]Clinical Trial Unit, Department of Clinical Research, University of Basel, University Hospital Basel, Basel, Switzerland

Corresponding Author:
Johannes Lorscheider, MD
Research Center for Clinical Neuroimmunology and Neuroscience Basel
University of Basel, University Hospital Basel
Spitalstrasse 2
Basel, 4056
Switzerland
Phone: 41 612652525
Email: johannes.lorscheider@usb.ch

Abstract

Background: Smartphones and their built-in sensors allow for measuring functions in disease-related domains through mobile tests. This could improve disease characterization and monitoring, and could potentially support treatment decisions for multiple sclerosis (MS), a multifaceted chronic neurological disease with highly variable clinical manifestations. Practice effects can complicate the interpretation of both improvement over time by potentially exaggerating treatment effects and stability by masking deterioration.

Objective: The aim of this study is to identify short-term learning and long-term practice effects in 6 active tests for cognition, dexterity, and mobility in user-scheduled, high-frequency smartphone-based testing.

Methods: We analyzed data from 264 people with self-declared MS with a minimum of 5 weeks of follow-up and at least 5 repetitions per test in the Floodlight Open study, a self-enrollment study accessible by smartphone owners from 16 countries. The collected data are openly available to scientists. Using regression and bounded growth mixed models, we characterized practice effects for the following tests: *electronic Symbol Digit Modalities Test (e-SDMT)* for cognition; *Finger Pinching* and *Draw a Shape* for dexterity; and *Two Minute Walk*, *U-Turn*, and *Static Balance* for mobility.

Results: Strong practice effects were found for *e-SDMT* (n=4824 trials), *Finger Pinching* (n=19,650), and *Draw a Shape* (n=19,019) with modeled boundary improvements of 40.8% (39.9%-41.6%), 86.2% (83.6%-88.7%), and 23.1% (20.9%-25.2%) over baseline, respectively. Half of the practice effect was reached after 11 repetitions for *e-SDMT*, 28 repetitions for *Finger Pinching*, and 17 repetitions for *Draw a Shape*; 90% was reached after 35, 94, and 56 repetitions, respectively. Although baseline performance levels were highly variable across participants, no significant differences between the short-term learning effects in low performers (5th and 25th percentile), median performers, and high performers (75th and 95th percentile) were found for *e-SDMT* up to the fifth trial (β=1.50-2.00). Only small differences were observed for *Finger Pinching* (β=1.25-2.5). For *U-Turn* (n=15,051) and *Static Balance* (n=16,797), only short-term learning effects could be observed, which ceased after a maximum of 5 trials. For *Two Minute Walk* (n=14,393), neither short-term learning nor long-term practice effects were observed.

Conclusions: Smartphone-based tests are promising for monitoring the disease trajectories of MS and other chronic neurological diseases. Our findings suggest that strong long-term practice effects in cognitive and dexterity functions have to be accounted for to identify disease-related changes in these domains, especially in the context of personalized health and in studies without a comparator arm. In contrast, changes in mobility may be more easily interpreted because of the absence of long-term practice effects, even though short-term learning effects might have to be considered.

(J Med Internet Res 2021;23(11):e30394) doi:10.2196/30394

KEYWORDS

multiple sclerosis; digital biomarkers; practice effects; learning effects; learning curves; nonlinear mixed models; quantile regression; information processing speed; symbol digit modalities test; smartphones; wearable electronic devices; mobile phones

Introduction

Background

Multiple sclerosis (MS) is a multifaceted and variable chronic autoimmune neurological disease affecting approximately 2.3 million people worldwide [1]. It is among the most common causes of nontraumatic disabilities in young adults [2].

MS progresses in different phases with highly variable speed and severity. To optimize treatment strategies, timely and precise monitoring of patients' disease status is essential. As MS affects multiple functional domains, a range of validated clinical tests are used: for cognition, the Symbol Digit Modalities Test (SDMT) measures mental processing speed and is highly established as a screening tool for cognitive impairment in MS [3,4]. The 9-hole peg test (9HPT) is routinely used to measure dexterity [5,6], and the timed 25-foot walk (T25FW) is used to measure mobility [7]. Usually, stable patients with MS have half-yearly to yearly clinical routine visits with neurologic examinations and regular magnetic resonance imaging (MRI), limiting insight into symptom fluctuations and reversible deficits [8].

Wearable technologies, such as smartphones and smartwatches, are expected to capture more representative data at a higher resolution not only in the patients' natural environments in MS but also in other neurological diseases such as Parkinson disease and Huntington disease [9-11]. Data can be collected passively during the patient's everyday activities (eg, capturing step counts, turn speed, or keyboard dexterity [9,12,13]) or actively during specific functional tests [14-16]. They can possibly improve both clinical trials by providing more sensitive outcome measures and clinical practice by allowing more personalized disease course monitoring [14,17].

Acknowledged difficulties in interpreting the results of repeated tests are learning and practice effects, especially in neuropsychology [18,19]. Without a comparator, it is difficult to disentangle whether longitudinal improvement constitutes remission, practice, or treatment effects. In the same light, disease progression and worsening of disability may be masked by practice effects when specific tests feign stability. These issues have been adequately addressed by control groups in randomized controlled trials [20]. For trials without control groups and for intraindividual comparisons—a cornerstone of personalized medicine—interpretation remains challenging. Furthermore, practice effects hamper test-retest reliability, which is illustrated by recommendations to discard the results from 3 prebaseline repetitions of the MS functional composite [21]. However, more recently, it has been suggested that person-specific learning curves can be used as new outcome measures, leveraging the information inherent in practice effects [14].

Objective

The aim of this analysis is to examine short-term learning and long-term practice effects in high-frequency smartphone-based tests representative of the assessment of 3 domains often affected by MS: cognition, dexterity, and mobility.

Methods

Study Data and Participant Selection

We used publicly available data from the *Floodlight Open* study, which collects smartphone-based test data from self-declared persons with MS with a number of different tests implemented in the Floodlight Open app [22]. The study is the successor of a small, closed feasibility study [9,17], and the data are openly available to researchers [23]. Currently, several phase 3 studies are using variations of the Floodlight app as part of their test batteries, for example, the CONSONANCE trial, a single-arm interventional trial evaluating ocrelizumab treatment in participants with progressive MS (NCT03523858) [17]. Recruitment for *Floodlight Open* started in April 2018, and some participants have been using the app continuously since then, amounting to more than 3 years of follow-up. However, most patients have only used the app for a very short time, leading to a strong right-skewness of the distribution of follow-up times. Among the 1147 patients who have performed at least one smartphone-based *e-SDMT* test in the period we examined, the median number of repetitions was 2, the IQR was 1-4 and the range was 1-119.

We included data up to and including July 31, 2021, and focused our analyses on the following 6 tests [9]: *e-SDMT* for cognition; *Finger Pinching* and *Draw a Shape* for dexterity; *Two Minute Walk*, *U-Turn*, and *Static Balance* for mobility. The dexterity tests have been shown to correlate with the 9HPT, the first 2 mobility tests with the T25FW and *Static Balance* with the Berg Balance Scale [9]. The Floodlight Open app allows performing *e-SDMT* up to a weekly frequency and all other tests up to a daily frequency, but the actual frequency was completely determined by the participant's choice. For dexterity tests, the left and right hands were alternated.

The *e-SDMT* consisted of consecutively tapping symbol-corresponding digits on a number pad on the smartphone screen as quickly as possible for 60 seconds. Thus, there was a dexterity component that may potentially introduce bias. Floodlight's *e-SDMT* included a second step termed *baseline*, simply showing digits instead of symbols, asking users to consecutively tap these digits on the same number pad for 15 seconds, without the symbol-association task. Using this second step by taking the quotient of the correct responses of the main test and the *baseline* potentially corrects for dexterity and reaction speed, representing only the true information processing speed.

Participants were selected for each test separately if at least 5 repetitions per test and at least 5 weeks between their first and

last repetitions were available. This yielded slightly different but largely overlapping subsets of participants for each test.

Statistical Analysis

Short-term Learning and Long-term Practice Effects

First, summary analyses were performed to investigate the mean scores of the first, fifth, and last trials of each test. We assumed that improvements up to the fifth score were more likely due to short-term learning effects, where participants learned to execute a test, and improvements from the fifth trial onward were more likely because of long-term practice effects. Naturally, these effects are intertwined, but using the fifth trial as the baseline was supported by Solari et al [24].

To examine group differences in baseline performances and potential short-term learning effects in low and high performers, linear quantile regression was performed on each test for the first 5 trials for the 5th, 25th, 50th, 75th, and 95th percentiles. Quantile regression *P* values were corrected with the Bonferroni method, and the 5 slopes were compared with an analysis of variance (ANOVA)–type test.

Long-term practice effects were assumed for tests with a significant mean difference from the fifth to the last score. The positive association of this difference with the number of repetitions (log-transformed to account for the strong right-skewness) adjusted for the potential confounders, age, first score, and fifth score, was considered as an additional indicator of long-term practice effects.

Long-term Learning Curve Analysis

For tests suggestive of long-term practice effects that meet the 2 abovementioned criteria, learning curve analysis was performed with 1 nonparametric and 3 parametric mixed effect models of increasing complexity, each modeling performance as a function of repetition, grouping by patient for cognition and mobility and by hand for dexterity. The performance of the 4 models was compared using both root mean squared error (RMSE) and the number of (effective) *df* used.

For the nonparametric model, smoothing splines calculated by generalized additive models were fitted to examine the unbiased shape of the potential learning curves, exhibiting different effective *df* per test [25,26].

For the parametric models, simple linear (*df*=4) and linear quadratic (*df*=5) mixed models were fitted, both using time and in addition the latter using time squared as fixed effects. As the third parametric model, we considered bounded growth mixed models (*df*=6) using the following formula:

$$y_{(t)} = \text{boundary} + (y_0 - \text{boundary})\, e^{-ct} \quad \textbf{(1)}$$

We treated boundary and baseline (y_0) as random effects, while we considered the growth constant *c* as a fixed effect.

Sensitivity Analyses

In addition to our main analysis on practice effects as a function of repetition with the selection criteria of a minimum of 5 weeks and 5 repetitions, we performed 3 additional sensitivity analyses: sensitivity analyses 1 and 3 were modeling practice effects as a function of weeks since the first test instead of the number of repetitions, and sensitivity analyses 2 and 3 were performed using stricter selection criteria of a minimum of 10 weeks and 10 repetitions (Table 1).

Table 1. Comparison of the main analysis with the 3 sensitivity analyses performed.

Criteria	Minimum of 5 weeks and 5 repetitions	Minimum of 10 weeks and 10 repetitions
Practice effects as a function of number of repetitions	Main analysis	Sensitivity analysis 2
Practice effects as a function of weeks since first test	Sensitivity analysis 1	Sensitivity analysis 3

All statistical analyses were performed using R 4.0.3 (R Foundation for Statistical Computing). Point estimates are accompanied by 95% CI in brackets, unless otherwise stated. *P* values were based on two-tailed *t* tests, unadjusted unless otherwise stated and considered significant if <.05. All analysis codes can be found on the web [27]. The data set used can be found on the web [28].

Results

Overview

Of the 1147 patients who performed at least one cognitive *e-SDMT*, 262 (22.8%) fulfilled our selection criteria of a minimum of 5 repetitions and 5 weeks between the first and last repetitions, accounting for 77.31% (4824/6240) of all performed *e-SDMT* tests. For *Finger Pinching* and *Draw a Shape*, 23.8% (264/1109) and 24% (259/1079) patients were selected, accounting for 87.14% (19,650/22,550) and 87.18% (19,019/21,816) of the performed tests, respectively. For *Two Minute Walk*, *U-Turn*, and *Static Balance*, 29.7% (171/575), 24.1% (217/901), and 24.34% (257/1056) patients were selected, representing 92.79% (14,393/15,512), 89.37% (15,051/16,841), and 88.41% (16,797/19,000) of the respective tests (Table 2). The minimum intertest interval was constrained by the app to 7 days for *e-SDMT*, 2 days for *Finger Pinching* and *Draw a Shape* for each hand, and 1 day for *Two Minute Walk*, *U-Turn*, *and Static Balance*, explaining the lower number of *e-SDMT* repetitions. However, participants had highly variable intertest intervals, making this an irregular time series, as indicated by each participant's median intertest interval and IQR. Table 2 shows the median of these statistics for all the selected participants.

Table 2. Characteristics of included patients with multiple sclerosis.

Domain	Cognition	Dexterity		Mobility		
	Electronic Symbol Digit Modalities Test	*Finger Pinching*	*Draw a Shape*	*Two Minute Walk*	*U-Turn*	*Static Balance*
Number of patients meeting selection criteria						
Total, N	1147	1109[a]	1079[b]	575	901	1056
Selected, n (%)	262 (22.8)	264 (23.8)	259 (24)	171 (29.7)	217 (24.1)	257 (24.3)
Number of tests performed by these patients						
Total, N	6240	22,550	21,816	15,512	16,841	19,000
Selected, n (%)	4824 (77.3)	19,650 (87.1)	19,019 (87.2)	14,393 (92.8)	15,051 (89.4)	16,797 (88.4)
Sex						
Total, N	262	499	484	171	217	257
Female, n (%)	184 (70.2)	353 (70.7)	345 (71.3)	123 (71.9)	155 (71.4)	181 (70.4)
Age (years), median (IQR; range)	50.2 (42.0-58.0; 20.0-79.0)	50.0 (41.8-58.0; 20.0-79.0)	49.6 (41.5-58.0; 20.0-79.0)	50.0 (41.6-58.1; 20.0-74.5)	49.6 (41.5-57.0; 20.0-79.0)	48.7 (41.1-57.0; 20.0-79.0)
Number of repetitions, median (IQR; range)	11 (7-18; 5-119)	17 (9-41.5; 5-416)	17 (9-41; 5-414)	30 (15-83.5; 5-827)	24 (11-69; 5-829)	24 (12-67; 5-828)
Median of intertest intervals (days), median (IQR; range)	7.9 (7.1-10.2; 6.7-87.1)	3.1 (2.1-5.4; 1.9-42.8)	3.3 (2.2-5.6; 1.9-77.6)	1.4 (1.0-3.0; 1.0-24.9)	1.8 (1.1-3.3; 1.0-25.4)	1.7 (1.1-3.1; 0.7-28.4)
Median of IQR of intertest intervals (days), median (IQR; range)	3.6 (1.0-9.9; 0.0-133.8)	3.3 (1.1-8.0; 0.0-198.1)	3.4 (1.1-8.8; 0.0-251.0)	2.0 (0.9-5.1; 0.1-39.2)	2.6 (0.9-6.9; 0.0-81.5)	2.3 (0.8-7.7; 0.0-93.0)
Number of weeks from the first to the last test, median (IQR; range)	18.3 (11.5-54.1; 5.0-164.7)	17.4 (10.9-49.4; 5.0-164.7)	17.9 (11.0-49.9; 5.0-164.7)	17.2 (11.1-48.2; 5.0-146.3)	16.3 (10.3-47.0; 5.0-146.3)	16.7 (10.3-47.0; 5.0-152.1)

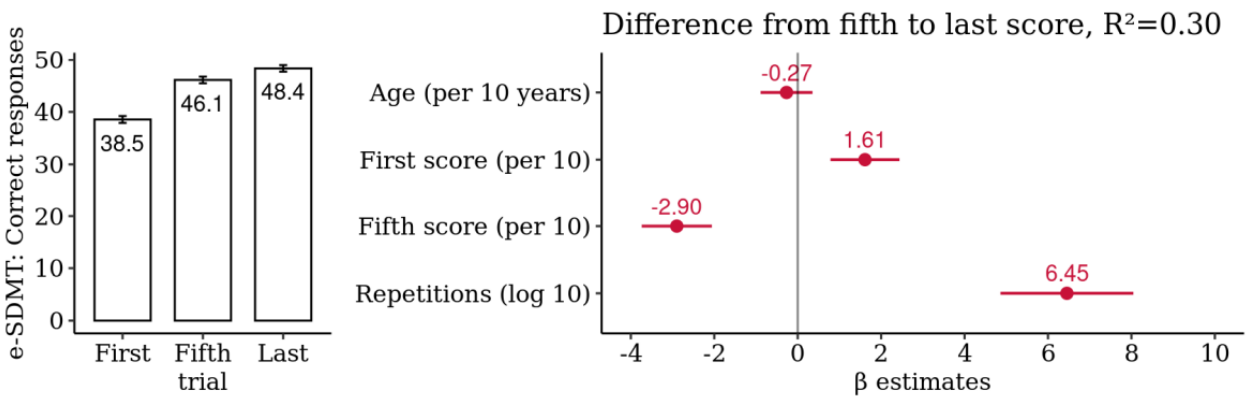
[a]With 26.19% (499/1905) of hands selected.

[b]With 26.23% (484/1845) of hands selected.

Cognition: e-SDMT

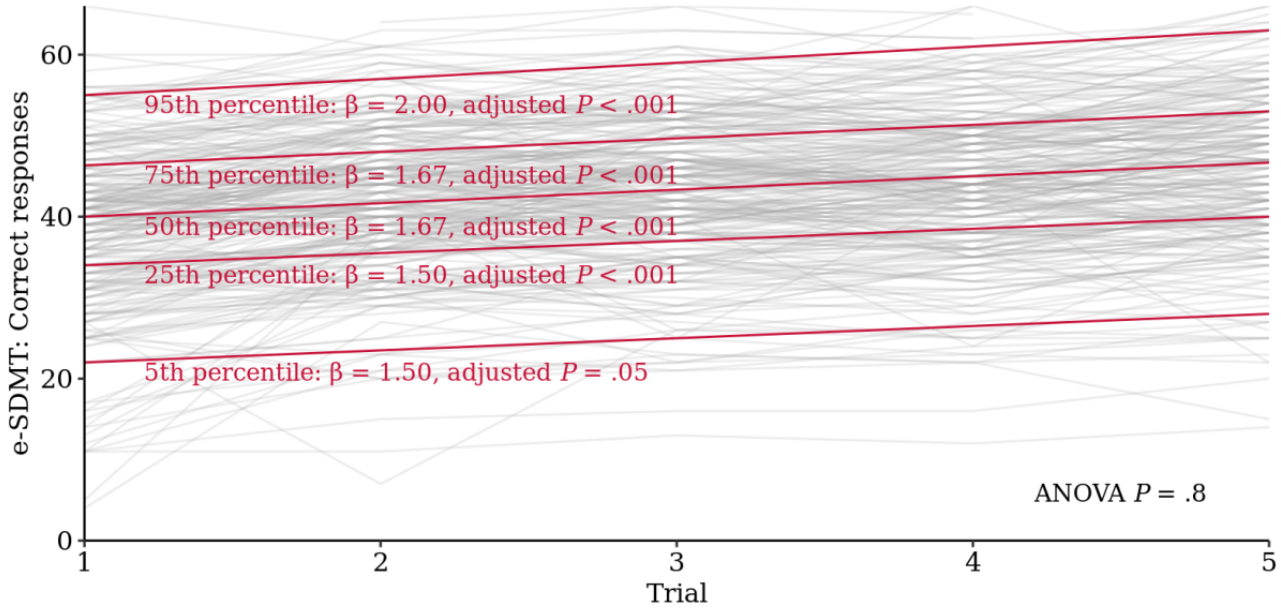
A summary analysis of the 262 selected patients yielded a mean difference from the first to last score of 9.8 correct responses, representing an average observed improvement of 25.4% (95% CI 23.1% to 27.8%) from the first score. Although the majority of this improvement (19.7%, 95% CI 17.5% to 21.9%) occurred up to the fifth score and can thus be considered a short-term learning effect, there was still a significant improvement from the fifth score onward of, on average, 5.7% (95% CI 4.1% to 7.4%), suggesting a long-term practice effect. A multivariate regression model of this difference yielded a significant association with the total number of repetitions, further supporting the long-term practice effects (Figure 1). Age was positively correlated with the number of repetitions performed (Pearson correlation coefficient, R=0.19; P=.003), but the first score was not (R=−0.03; P=.70; Multimedia Appendix 1).

Figure 1. Patient-level summary analysis for the *electronic Symbol Digit Modalities Test*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to last score with age, first and fifth score, and the log-transformed number of repetitions (n=262 patients). *e-SDMT*: *electronic Symbol Digit Modalities Test.*



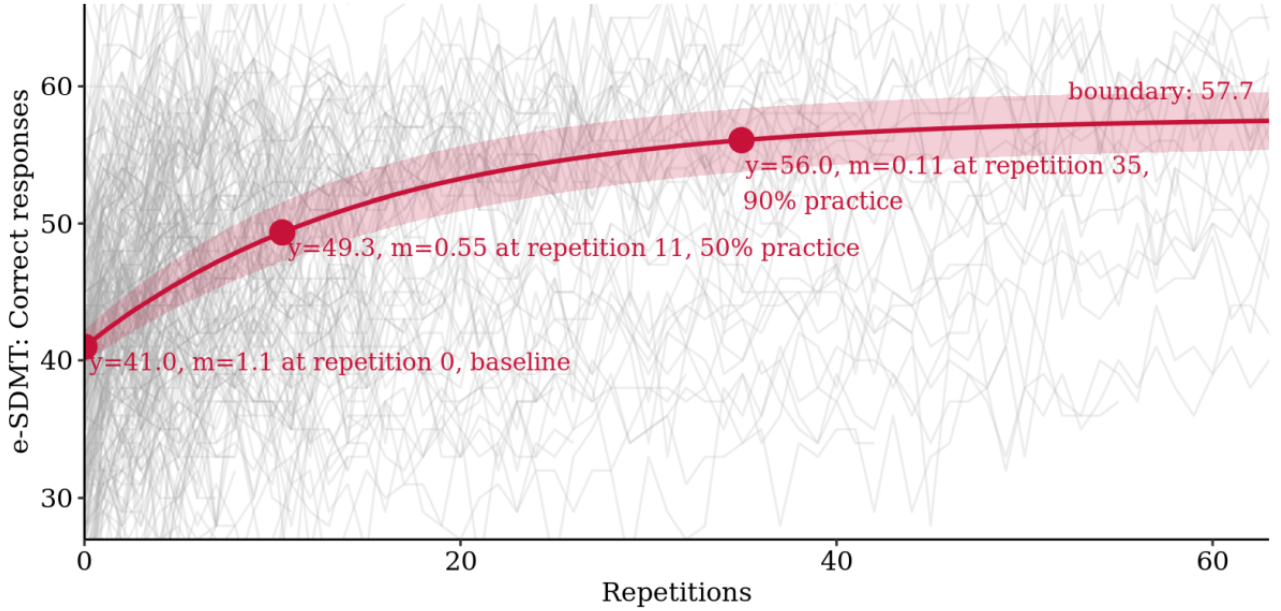
When comparing performances by 5th, 25th, 50th, 75th, and 95th percentile groups up to the fifth trial with quantile regression, baseline performances were normally distributed with intercept estimates of 22.0 (19.1-24.9), 34.0 (32.7-35.3), 40.0 (38.8-41.2), 46.3 (45.1-47.6), and 55.0 (53.4-56.6), respectively. The ANOVA-type test for all 5 slopes (β=1.5-2.0) did not suggest that short-term learning rates for these groups differed significantly (*P*=.80; Figure 2).

Figure 2. Linear quantile regression for the *electronic Symbol Digit Modalities Test* of short-term learning effects up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression *P* values are Bonferroni-adjusted (n=1310 tests). ANOVA: analysis of variance; *e-SDMT*: *electronic Symbol Digit Modalities Test*.



The long-term learning curve analysis showed that the bounded growth model fit the data best with an RMSE of 3.3 correct responses, followed by 3.6 for the smoothing spline, 3.8 for the quadratic, and 4.0 for the linear model (Multimedia Appendix 2). Strong boundary practice effects were found with baseline estimates of on average 41.0 (95% CI 39.8 to 42.2) correct responses and boundary estimates of 57.7 (95% CI 55.7 to 59.8) correct responses, leading to an average improvement over baseline of 40.8% (95% CI 39.9% to 41.6%). Half of the practice effect was reached after 11 repetitions and 90% after 35 repetitions (Figure 3).

Figure 3. Learning curve analysis for the *electronic Symbol Digit Modalities Test*: bounded growth mixed model of practice effects with 95% CI band and baseline, 50% and 90% practice points marked (m=slope of tangent; n=4824 tests). *e-SDMT*: *electronic Symbol Digit Modalities Test*.



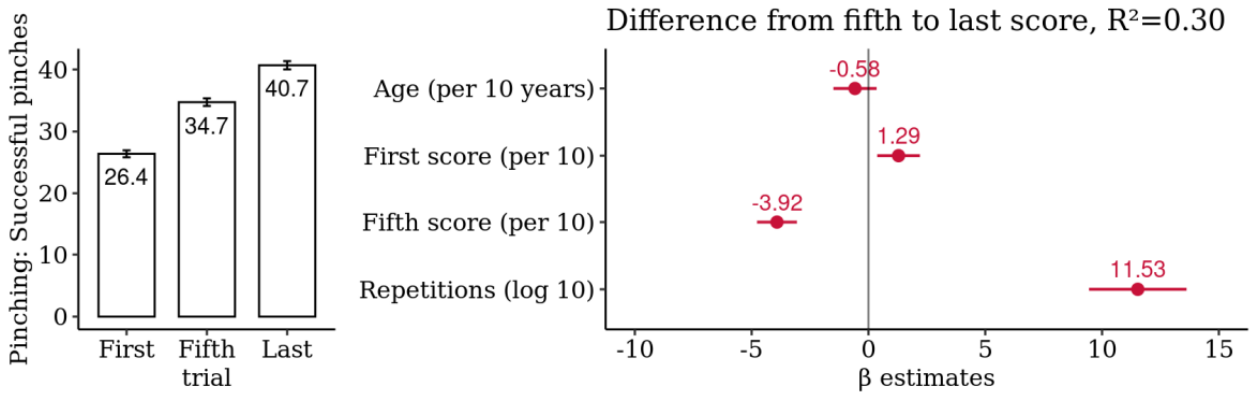
For *e-SDMT* corrected for dexterity and reaction speed, Spearman correlation of all 6190 corrected scores with their uncorrected counterpart yielded ρ=0.55. The resulting practice effects were very similar to the uncorrected *e-SDMT*, with an observed improvement from the first to last score of 19% (95% CI 16.1% to 22%), consisting of 12.4% (95% CI 9.8% to 15%)

from the first to fifth score, and 6.6% (95% CI 4.6% to 8.7%) from the fifth to last score. As with the uncorrected *e-SDMT*, no significant differences in short-term learning rates were found between low, median, and high performers (ANOVA-type *P*=.30). Boundary long-term improvements were 23.5% (95% CI 23% to 24%), with half of the practice effect reached after 7 repetitions and 90% practice effect reached after 23 repetitions (Multimedia Appendices 3-6).

Dexterity: Finger Pinching

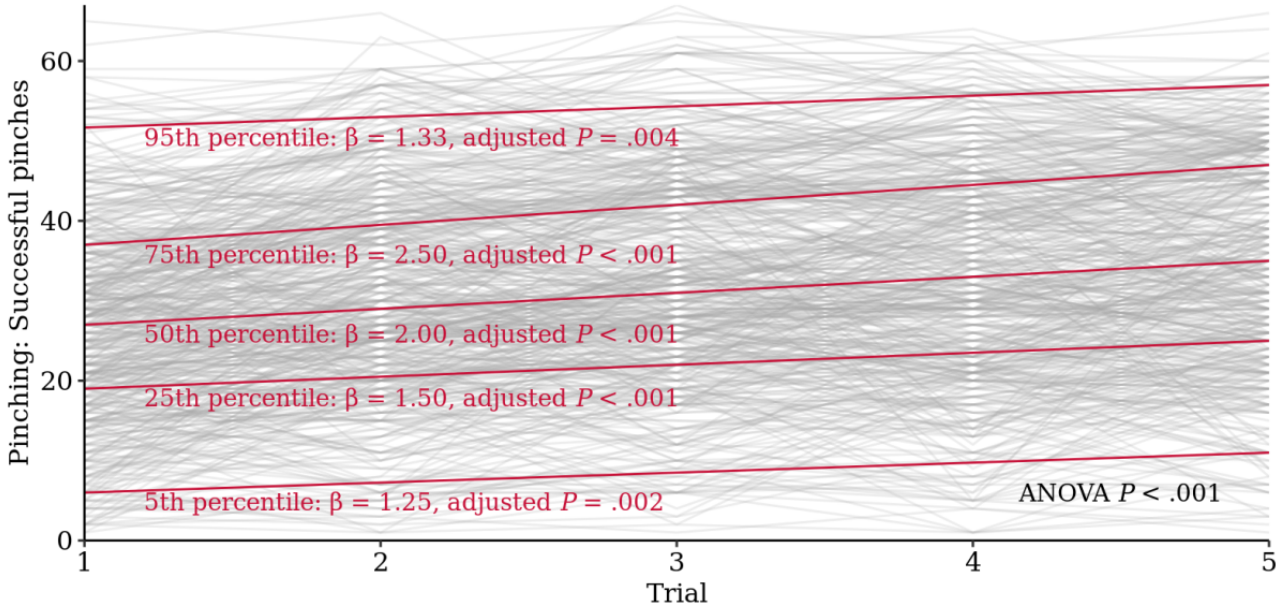
A summary analysis of the 499 selected hands yielded a mean difference from the first to last score of 14.3 successful pinches, representing an average observed improvement of 54.2% (95% CI 49.3% to 59.1%) over the first score. Similar to the findings on the *e-SDMT*, the majority of this improvement (31.5%, 95% CI 27.5% to 35.4%) occurred up to the fifth score, compatible with a short-term learning effect. However, the remaining improvement of 22.7% (95% CI 18.6% to 26.8%) occurred after the fifth trial. This improvement was significantly associated with the total number of repetitions, indicating a strong long-term practice effect (Figure 4). Age was positively correlated with the number of repetitions performed (*R*=0.21; *P*<.001) but the first score was not (*R*=−0.06; *P*=.20; Multimedia Appendix 7).

Figure 4. Hand-level summary analysis for *Finger Pinching*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to last score with age, first and fifth score, and the log-transformed number of repetitions (n=499 hands).



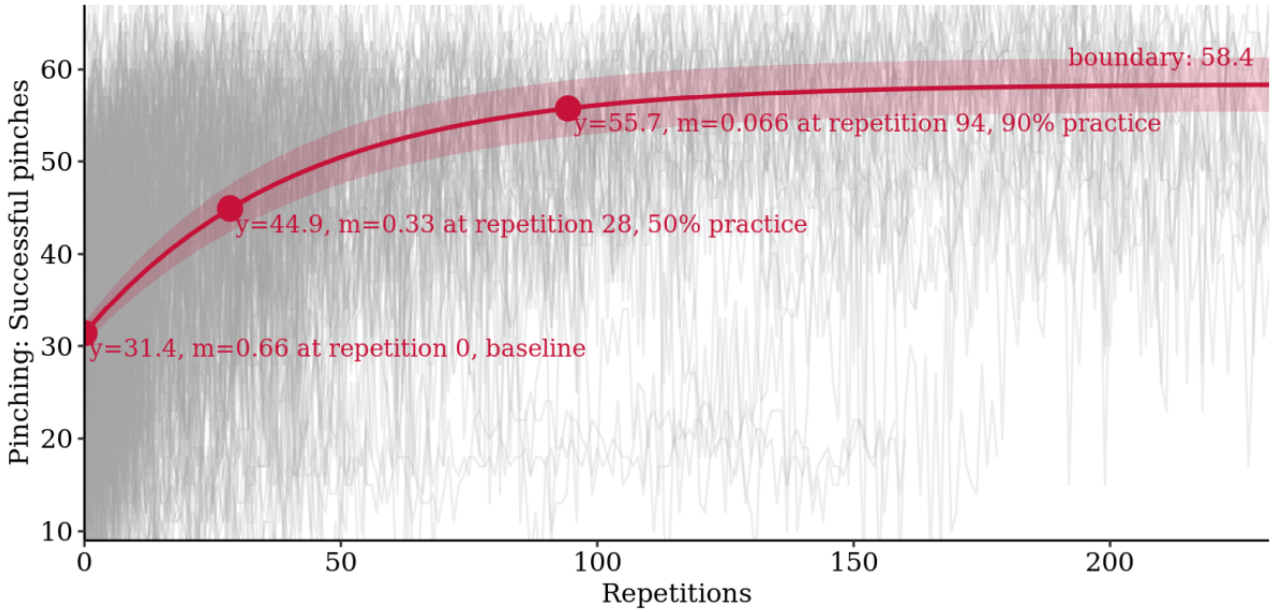
Baseline performances were normally distributed with intercept estimates of 6.0 (95% CI 4.5 to 7.5) for the fifth percentile, 19.0 (95% CI 17.8 to 20.2) for the 25th, 27.0 (95% CI 25.8 to 28.2) for median performers, 37.0 (95% CI 35.6 to 38.4) for the 75th, and 51.7 (95% CI 49.6 to 53.8) for the 95th percentile with quantile regression. The β coefficients for short-term learning up to the fifth trial were the highest for the 75th percentile and median performers with 2.50 (95% CI 1.96 to 3.04) and 2.00 (95% CI 1.45 to 2.55) additional successful pinches per repetition, lower for the 25th percentile (1.50, 95% CI 1.00 to 2.00) and the lowest for the 5th and 95th percentiles (1.25, 95% CI 0.57 to 1.93, and 1.33, 95% CI 0.56 to 2.11, respectively). These differences in slopes between performance levels were significant (ANOVA-type *P*<.001; Figure 5).

Figure 5. Linear quantile regression for *Finger Pinching* of short-term learning effects up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression *P* values are Bonferroni-adjusted (n=2495 tests). ANOVA: analysis of variance.



Long-term learning curve analysis again showed that the bounded growth model fit the data best with an RMSE of 6.8 successful pinches, followed by 7.5 for the smoothing spline, 7.9 for the quadratic, and 8.1 for the linear model (Multimedia Appendix 8). Strong boundary practice effects were found with baseline estimates of, on average, 31.4 (95% CI 30.2 to 32.5) and boundary estimates of 58.4 (95% CI 55.5 to 61.4) successful pinches, leading to an average improvement over baseline of 86.2% (95% CI 83.6% to 88.7%). Half of the practice effect was reached after 28 repetitions and 90% after 94 repetitions (Figure 6).

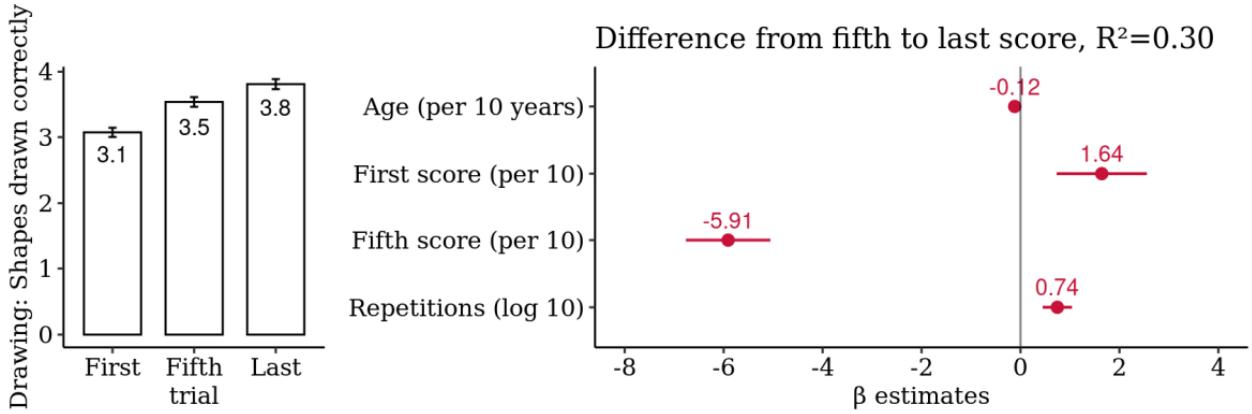
Figure 6. Learning curve analysis for *Finger Pinching*: bounded growth mixed model of practice effects with 95% CI band and baseline, 50%, and 90% practice points marked (m=slope of tangent; n=19,650 tests).



Dexterity: Draw a Shape

A summary analysis of the 484 selected hands yielded a mean improvement in the number of shapes drawn correctly from the first to last score of 23.9% (95% CI 18.3% to 29.5%), from the first to fifth score of 15.1% (95% CI 9.8% to 20.3%), and from the fifth to last score of 8.8% (95% CI 3.8% to 13.8%). This difference was significantly associated with the total number of repetitions, suggesting a long-term practice effect (Figure 7). Age was positively correlated with the number of repetitions performed (R=0.22; P<.001) but the first score was not (R=−0.08; P=.09; Multimedia Appendix 9).

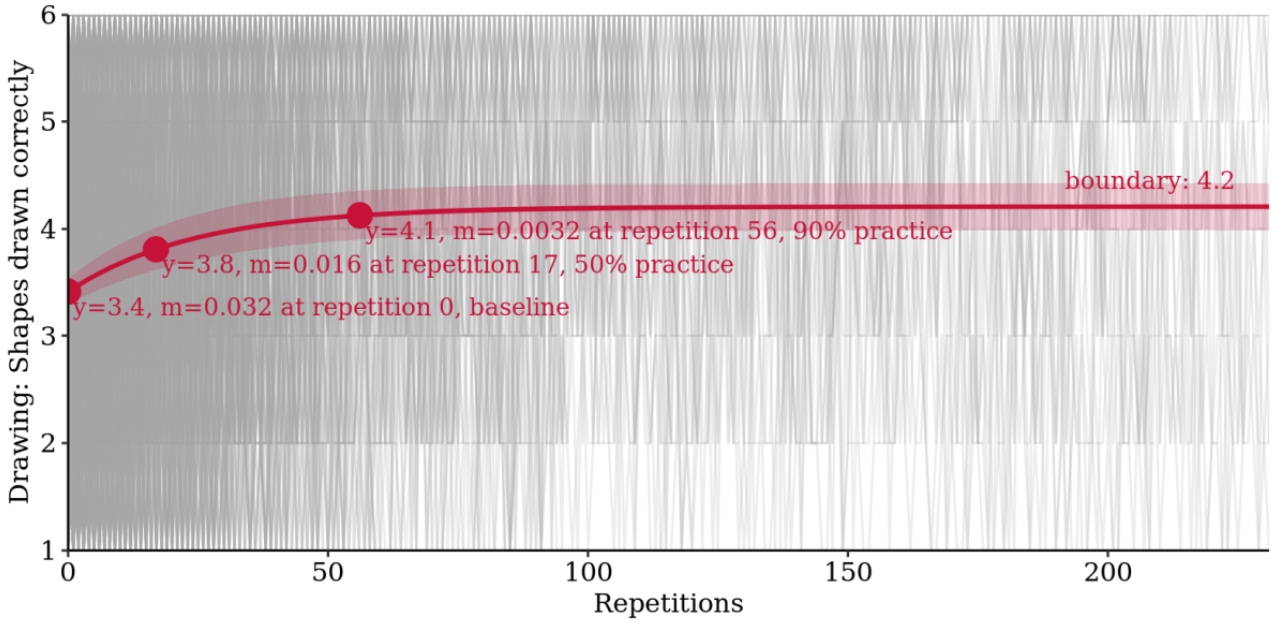
Figure 7. Hand-level summary analysis for *Draw a Shape*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to the last score with age, first and fifth score, and the log-transformed number of repetitions (n=484 hands).



Intercept estimates for baseline performances were 1 shape drawn correctly for the 5th percentile, 2 for the 25th percentile, 3 for the median performers, 5 for the 75th percentile, and 6 for the 95th percentile with quantile regression. In this analysis, only median performers showed a significant short-term learning rate up to the fifth trial (Multimedia Appendix 10).

The long-term learning curve analysis again showed that bounded growth models fit the data best with an RMSE of 1.02 shape drawn correctly, followed by 1.06 for the smoothing spline, 1.07 for the quadratic, and 1.08 for the linear model (Multimedia Appendix 11). Boundary practice effects were found with an average improvement over baseline of 23.1% (95% CI 20.9% to 25.2%), reaching half of the practice effect after 17 repetitions and 90% after 56 repetitions (Figure 8).

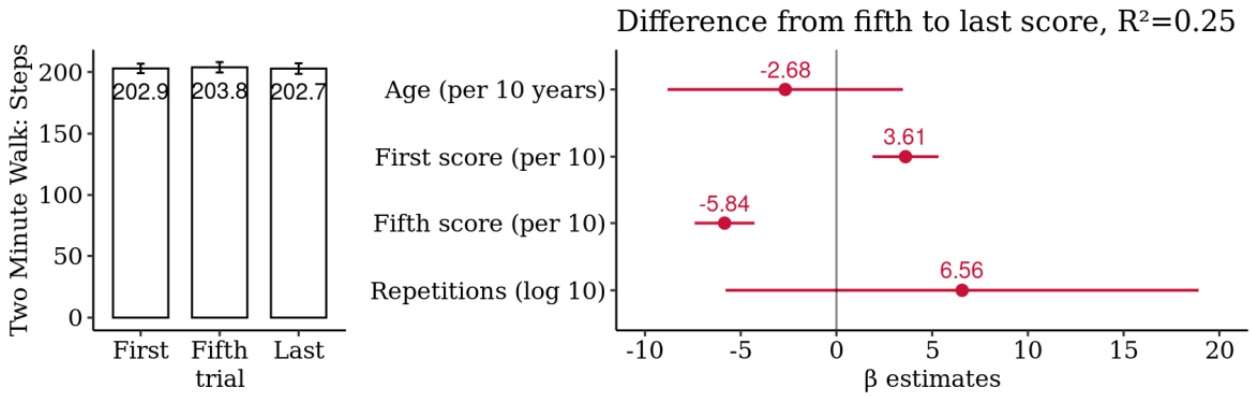
Figure 8. Learning curve analysis for *Draw a Shape*: bounded growth mixed model of practice effects with 95% CI band and baseline, 50%, and 90% practice points marked (m=slope of tangent; n=19,019 tests).



Mobility: Two Minute Walk

A summary analysis of the 171 selected patients yielded no significant difference between the first, fifth, and last scores with a mean difference from the fifth to last score of 1.4 (95% CI −5.2 to 7.9) steps. This difference was also not associated with the total number of repetitions performed (Figure 9 and Multimedia Appendix 12).

Figure 9. Patient-level summary analysis for *Two Minute Walk*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to the last score with age, first and fifth score, and the log-transformed number of repetitions (n=171 patients).

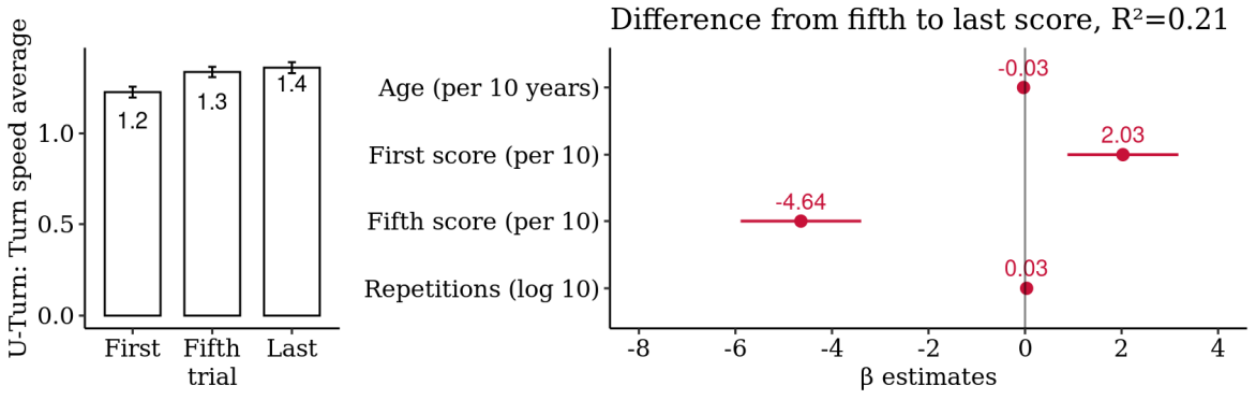


The distribution of baseline performance was left-skewed with performers in the 5th percentile achieving, on average, 87.0 (95% CI 52.2 to 121.8) steps; in the 25th percentile, 181.0 (95% CI 170.7 to 191.3) steps; on median, 219.0 (95% CI 212.6 to 225.4) steps; in the 75th percentile, 236.0 (95% CI 230.6 to 241.4) steps; and in the 95th percentile, 260.0 (95% CI 255.2 to 264.8) steps. No significant slopes up to the fifth trial could be observed (Multimedia Appendix 13).

Mobility: U-Turn

A summary analysis of the 217 selected patients yielded a significant improvement from the first to last score with a mean difference in turn speed average of 0.13 rad/s, representing an average observed difference of 11.0% (95% CI 5.7% to 16.2%) over the first score. However, the majority of this difference occurred up to the fifth score (9%, 95% CI 3.7% to 14.3%), and the remaining difference from the fifth to last score (1.9%, 95% CI −2.3% to 6.1%) was neither significant nor associated with the total number of repetitions performed (Figure 10 and Multimedia Appendix 14).

Figure 10. Patient-level summary analysis for *U-Turn*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to the last score with age, first and fifth score, and the log-transformed number of repetitions (n=217 patients).

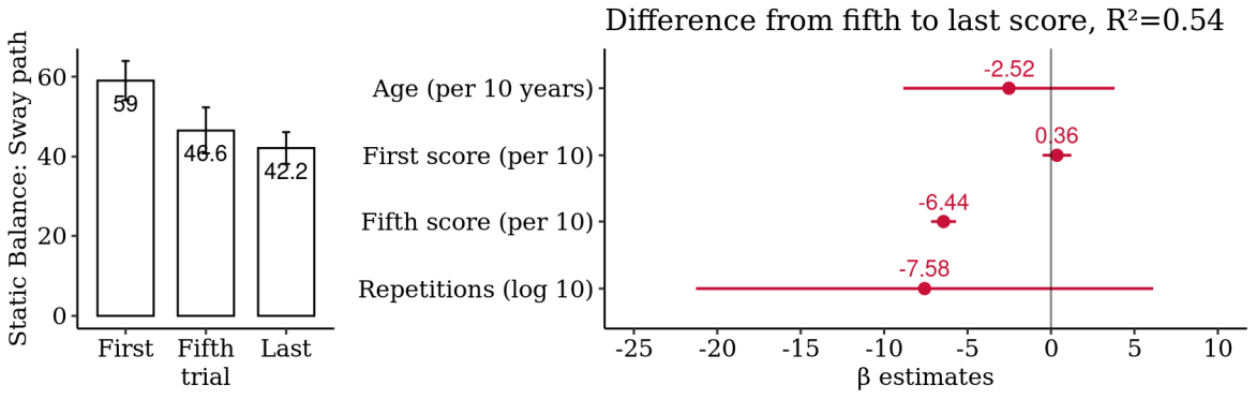


Baseline performances estimated with quantile regression were normally distributed with 0.5 rad/s (95% CI 0.5 to 0.6) for the 5th percentile, 0.9 rad/s (95% CI 0.9 to 1.0) for the 25th percentile, 1.3 rad/s (95% CI 1.2 to 1.3) for median performers, 1.5 rad/s (95% CI 1.5 to 1.6) for the 75th percentile, and 2.0 rad/s (95% CI 1.9 to 2.1) for the 95th percentile groups. Only the slope of the 25th percentile group was significant in this analysis up to the fifth trial (β=.04; 95% CI 0.02 to 0.06), and the difference in slopes was not significant in the ANOVA-type test (*P*=.40; Multimedia Appendix 15).

Mobility: Static Balance

A summary analysis of the 257 selected patients yielded a significant difference from the first to last score, with a mean difference in sway path of −16.9 m/s². This is the only test in which fewer numbers are better. Thus, the average observed improvement was −28.6% (95% CI −48.6% to −8.5%) over the first score. However, the majority of this improvement occurred up to the fifth score (−21.1%, 95% CI −45% to −2.8%), and the remaining difference from the fifth to last score (−7.5%, 95% CI −24.1% to 9.2%) was neither significant nor associated with the total number of repetitions performed (Figure 11 and Multimedia Appendix 16).

Figure 11. Patient-level summary analysis for *Static Balance*: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to the last score with age, first and fifth score, and the log-transformed number of repetitions (n=257 patients).



Baseline performance estimates were strongly right-skewed with 5.7 m/s² (95% CI 4.3 to 7.0) for the 5th percentile, 11.7 m/s² (95% CI 9.4 to 14.0) for the 25th percentile, 23.8 m/s² (95% CI 20.7 to 26.8) for the median performers, 68.0 m/s² (95% CI 55.7 to 80.3) for the 75th percentile, and 260.2 m/s² (95% CI 200.4 to 320.0) for the 95th percentile. In this test, the 5th and 25th percentiles are the top performing groups, and their quantile regression slopes up to the fifth trial are not significant. However, the significant negative slopes of median (β=−2.32; 95% CI −3.47 to −1.17), 75th percentile (β=−8.46; 95% CI −12.21 to −4.72), and 95th percentile (β=−27.25; 95% CI −46.30 to −8.20) performers were increasingly steep, and the overall ANOVA-type difference test yielded *P*<.001(Multimedia Appendix 17).

Sensitivity Analyses

The results of the sensitivity analyses were in line with the results of the main analysis. Sensitivity analysis 2, which used stricter inclusion criteria with a minimum of 10 weeks and 10 repetitions, was overall very similar with expected further increases in mean improvement from the fifth to last score (mean improvement 5.7% for main analysis vs 9.1% for sensitivity analysis 2 for *e-SDMT*; 22.7% for main analysis vs 35.0% for sensitivity analysis 2 for *Finger Pinching*; and 8.8% for main analysis vs 13.3% for sensitivity analysis 2 for *Draw a Shape*) and a slight decrease in average boundary increase in performance (average boundary increase 40.8% for main analysis vs 34.3% for sensitivity analysis 2 for *e-SDMT*; 86.2% for main analysis vs 73.8% for sensitivity analysis 2 for *Finger*

Pinching; and 23.1% for main analysis vs 19.8% for sensitivity analysis 2 for *Draw a Shape*). Sensitivity analyses 1 and 3, which modeled practice effects as a function of weeks since the first test instead of the number of repetitions, also supported the main findings. However, the association of the maximum number of weeks from the first to the last tests with the difference from the fifth to last score was generally lower and so was the average increase in performance (Multimedia Appendix 18).

Discussion

Principal Findings

Strong long-term practice effects were found for *e-SDMT*, *Finger Pinching*, and *Draw a Shape*, with mean observed improvements of 25.4%, 54.2%, and 23.9% from the first to last score, respectively. Of these, 5.7%, 22.7%, and 8.8% occurred from the fifth score onward. However, the number of repetitions differed widely among participants with a range of 5-119 repetitions for *e-SDMT* (median 11), 5-416 for *Finger Pinching* (median 17), and 5-414 for *Draw a Shape* (median 17), introducing bias. To estimate boundary practice effects independent of the number of repetitions in our sample, we modeled learning curves with bounded growth models, a subtype of nonlinear mixed models. This approach yielded boundary improvements over the baseline of 40.8% for *e-SDMT*, 86.2% for *Finger Pinching*, and 23.1% for *Draw a Shape*. Interestingly, the practice effect seemed to last longer for the dexterity tests *Finger Pinching* and *Draw a Shape*, reaching half of the practice effect after repetition 28 and 17, respectively, compared to repetition 11 for *e-SDMT*.

These practice effects likely include both short-term learning effects, where patients become acquainted with the tests, and long-term practice effects. We believe these effects have not only different origins, time scales, and magnitudes but also different implications for the use of digital assessments in clinical studies and clinical practice. Short-term learning effects can be addressed by ensuring that participants have sufficient training before the observational period; long-term practice effects constitute a significant challenge for all applications beyond trials with a comparator arm. Although these effects are impossible to untangle in an unsupervised setting like this, we considered improvements up to the fifth trial to be more likely due to short-term learning and improvements afterward more likely because of long-term practice effects, based on the recommendation to use the fifth trial of the 9HPT as baseline [24].

For *U-Turn* and *Static Balance*, only short-term learning effects could be observed, ceasing after a maximum of 5 repetitions. Interestingly, for *Static Balance*, these short-term learning effects were not present in those with high baseline performance and were most pronounced in those with low baseline performance, potentially highlighting that the test instructions were not clear from the beginning. For *Two Minute Walk*, neither short-term learning nor long-term practice effects were observed.

For *e-SDMT*, quantile regression analysis suggested that the short-term learning rate was independent of the baseline performance. However, for *Finger Pinching*, median and high performers improved significantly faster than low and highest performers, with the learning rate decreasing toward the extremes. One can hypothesize that low performers might be more physically disabled, preventing them from improving as quickly as the median performers. On the other hand, the highest performers might reach their boundary sooner, leaving less room for improvement.

The 3 sensitivity analyses confirmed our main findings. However, for sensitivity analyses 1 and 3, which modeled practice effects as a function of weeks since the first test instead of the number of repetitions, the effect sizes were smaller. We believe this is caused by the irregular nature of these time-series data, as the intertest intervals differed widely, highlighting a complication in user-scheduled testing (Table 2).

Comparison With Previous Work

Overview

Only a few studies have examined practice effects in smartphone-based tests for patients with MS. Bove et al [14] analyzed the data from 38 patients, 22 of whom completed the planned study period of 12 months. They found strong practice effects for both their custom-made cognitive tests (digital adaptations of the trail-making test, the n-back test, a verbal fluency test, and an attention test), and a digital adaptation of the 9HPT. Interestingly, they suggest using person-specific learning curves quantified by binary spline inflection point analysis as a potential outcome measure [14].

In addition, Liao et al [29] recently reported significant practice effects for information processing speed and manual dexterity but not for walking speed in a tablet-based test battery called *MS Performance Test*, broadly confirming our results. However, they only analyzed 2-5 repetitions per patient and per test and could thus not examine long-term practice effects. Interestingly, they found that younger age was associated with larger practice effects, whereas we observed the opposite for *e-SDMT* and *Finger Pinching* (Multimedia Appendices 1 and 7), potentially highlighting differences between their low-frequency and our high-frequency testing.

Cognition: SDMT

Practice effects are well known for SDMT in both healthy controls and patients with MS, although the effect sizes reported were highly variable. Morrow et al [30] studied 660 natalizumab-treated patients with MS with a total of 13 repetitions of the oral SDMT over 48 weeks with average baseline scores of 46.8 (SD 15.3) correct responses and average final scores of 62.2 (SD 18.1) at week 48, resulting in an average improvement of 32.9% over baseline. Although the improvement was most pronounced over the first 3 repetitions, there was no obvious boundary [30].

In contrast, Benedict et al [31] found only minimal practice effects in 76 patients with MS with a total of 6 repetitions of oral SDMT over 5 months. Average baseline scores of 49.8 (SD 12.4) correct responses and average final scores of 52.5 (SD 14.3) at month 5, representing an improvement of only 5.4% over baseline, were found. However, their 25 healthy controls

improved from 62.0 (SD 11.3) to 71.4 (SD 13.2), representing a practice effect of 15.1% [31].

Roar et al [32] examined practice effects in 80 natalizumab-treated patients with MS with up to 31 repetitions over 30 months and reported improvements of roughly 25% over baseline, on average, with the rate of improvement slowing down after 6 months. Baseline performance and relative improvement were worse for the more severely affected patients with MS [32]. Interestingly, rearrangement of the SDMT symbol key resulted in a return to baseline performance, suggesting that the practice effect could be attributed to key memorization and that no generalizable learning or improvement of processing speed occurred [32].

Indeed, all of the traditional paper and pencil SDMT versions have the limitation of a fixed key, which is why Benedict et al [33] recommend the use of equivalent forms with alternate keys to mitigate practice effects. The smartphone-based *e-SDMT* version examined in this study has changing keys, thus emulating the process of alternate forms. Key changes are not truly random, and the subtlety of the original SDMT, in which the first 26 items only use the first 6 symbols in the key, is preserved [34].

With our result of an average boundary improvement over baseline of 40.8%, we can show that with weekly testing, practice effects for SDMT are likely to be stronger than with monthly testing, as performed by the abovementioned studies, and at least partly independent of the key.

As a limitation, the smartphone-based test in this study was not oral but based on touching a number pad, potentially biasing the results by dexterity problems (and dexterous practice effects). However, our analysis of *e-SDMT* corrected for dexterity and reaction speed by using Floodlight's *baseline* showed very similar results, suggesting that this is not a major issue, potentially because patients are free to use their preferred hand for this test.

Dexterity: 9HPT

Practice effects often become apparent in the examination of test-retest, intrarater, and interrater reliability. In this way, Cohen et al [35] found evident practice effects for the 9HPT first in 10 and later in 436 secondary progressive patients with MS over 4-6 repetitions in up to 4 weeks, which stabilized after 3 repetitions [21].

Solari et al [24] found even stronger practice effects in 32 patients with MS with 6 repetitions in 24 hours, which stabilized after 4 repetitions. As a consequence, they recommend performing 9HPT 4 times before baseline in any study to mitigate practice effects [24].

The smartphone-based *Finger Pinching* and, to a slightly lesser extent, also the *Draw a Shape* tests we examined seem to have much stronger and longer-lasting practice effects than the 9HPT. It can be speculated that high-frequency testing (ie, sustained daily practice over multiple months) maximizes the long-term practice effect.

Mobility: T25FW

No practice effects were found for T25FW, which was examined alongside 9HPT in the abovementioned studies [21,24,35]. This result is in line with our finding of no long-term practice effect in the smartphone-based mobility tests, *Two Minute Walk* and *U-Turn*, which have been validated with the T25FW [9]. However, the outcome reported for *Two Minute Walk* in this study (ie, number of steps) is a quantitative gait parameter and thus representative of endurance, unlike the more qualitative gait parameter *step power* used by Montalban et al [9] for validation against T25FW.

Limitations

MS diagnoses of study participants were self-declared, and there was no confirmation or assessment by health professionals. In addition, no clinical information was available for the participants to compare with their performance in digital tests. Differences caused by disease duration, severity, or treatment could not be analyzed.

In addition, we observed a high variability of results, which is most likely partly due to biomedical day-to-day fluctuations and partly due to circumstantial and technical noise, for example, caused by interrupted test performance or sensor error. However, it is impossible to determine these effects using the present data set.

Finally, these time-series data are highly irregular and have strong right-skewness. Our models expect data missing at random. We found no evidence that baseline performance influenced adherence and the number of repetitions, but age was found to be a confounder for all domains. Interestingly, older people tended to perform more repetitions than younger people (R=0.19 for *e-SDMT*; R=0.21 for *Finger Pinching*; and R=0.22 for *Draw a Shape*; Multimedia Appendices 1, 7, and 9). One can hypothesize that this is because older patients with MS tend to be more severely affected and thus might have higher intrinsic motivation. Another possible explanation is that younger people might have more competing time commitments, for example, because of their occupation or family. As age was associated with larger observed improvements from the fifth to last score for *e-SDMT* and *Finger Pinching* (both R=0.16; Multimedia Appendices 1 and 7), this confounder might lead to slight overestimation of the practice effects. However, in the multivariate models, age was not a significant confounder (Figures 1 and 4). Nevertheless, there might be unobserved confounders that differ between those participants who quit early and those who stayed engaged for a long time, which we aimed to mitigate by sensitivity analyses 2 and 3.

Conclusions

In summary, we analyzed the practice effects in 6 active smartphone-based tests for cognition, dexterity, and mobility performed at high frequencies. Smartphone-based tests promise to help monitor MS disease trajectories, and there are currently multiple initiatives in development [22,36-39]. Our results suggest that strong long-term practice effects in cognitive and dexterity tests must be accounted for to identify possible disease-related changes in these domains, lasting for more than 35 repetitions for *e-SDMT*, 94 for *Finger Pinching*, and 56 for

Draw a Shape. This is important for the interpretation of these tests in the context of personalized health and in studies with no comparator arm. On the other hand, the lack of long-term practice effects in mobility tests simplifies their interpretation, even though short-term learning effects might have to be considered.

Acknowledgments

The authors thank all the participants of the Floodlight Open study for their participation and Roche for making the data set publicly available.

Authors' Contributions

TW designed the study, performed the analysis, interpreted the data, and drafted and revised the manuscript. JL designed the study, interpreted the data, and revised the manuscript. TW and JL had full access to the available source data and guaranteed the integrity of the analysis. SP, AW, LK, and YN helped interpret the data and revised the manuscript for important intellectual content.

Conflicts of Interest

The research activities of the Research Center for Neuroimmunology and Neuroscience Basel, the affiliation of TW, SP, LK, YN, and JL, are supported by the University Hospital Basel, the University of Basel, and by grants from Novartis and Roche. One of the main projects of Research Center for Neuroimmunology and Neuroscience Basel is the development of a new comprehensive MS digital solution. TW and SP report no further conflicts of interest. AW reports no conflicts of interest. The University Hospital Basel, as the employer of LK, has received and dedicated to research support fees for board membership, consultancy or speaking, or grants in the past 3 years from Actelion, Bayer, Biogen Idec, CSL Behring, Eli Lilly EU, Genmab, GeNeuro SA, Janssen, Merck Serono, Novartis, Roche, Santhera, Sanofi-Aventis, Teva, European Union, Innosuisse, Roche Research Foundation, Swiss MS Society, and Swiss National Research Foundation. The University Hospital Basel, as the employer of YN, has received financial support for lectures from Teva and Celgene and grant support from Innosuisse (Swiss Innovation Agency). JL received research support from Innosuisse, Biogen, and Novartis; he received speaker honoraria and/or served on advisory boards for Biogen, Novartis, Roche, and Teva.

Multimedia Appendix 1
Pairwise associations for the <italic>electronic Symbol Digit Modalities Test</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=262 patients).
[PNG File , 362 KB - jmir_v23i11e30394_app1.png]

Multimedia Appendix 2
Model selection for <italic>electronic Symbol Digit Modalities Test</italic>: Comparison of 4 mixed models of increasing complexity (parametric linear, quadratic and bounded growth models and the non–parametric smoothing spline model) by root mean squared error (RMSE) and (effective) degrees of freedom (eDF) used.
[PNG File , 343 KB - jmir_v23i11e30394_app2.png]

Multimedia Appendix 3
Comparison of correct responses of the <italic>electronic Symbol Digit Modalities Test</italic> (x-axis) and of the <italic>electronic Symbol Digit Modalities Test</italic> divided by baseline (y-axis) to correct for dexterity and reaction speed (n=6190 tests).
[PNG File , 220 KB - jmir_v23i11e30394_app3.png]

Multimedia Appendix 4
Patient-level summary analysis for the <italic>electronic Symbol Digit Modalities Test</italic> corrected for dexterity and reaction speed: comparison of the first, fifth, and last score. Multivariate association of the difference from the fifth to the last score with age, first and fifth score, and the log-transformed number of repetitions (n=262 patients).
[PNG File , 74 KB - jmir_v23i11e30394_app4.png]

Multimedia Appendix 5
Linear quantile regression for the <italic>electronic Symbol Digit Modalities Test</italic> (corrected for dexterity and reaction speed) of short-term learning effects up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression <italic>P</italic> values are Bonferroni-adjusted (n=1310 tests).

[PNG File , 268 KB - jmir_v23i11e30394_app5.png]

Multimedia Appendix 6
Learning curve analysis for the <italic>electronic Symbol Digit Modalities Test</italic> corrected for dexterity and reaction speed: bounded growth mixed model of practice effects with 95% CI band and baseline, 50%, and 90% practice points marked (m=slope of tangent; n=4801 tests).
[PNG File , 375 KB - jmir_v23i11e30394_app6.png]

Multimedia Appendix 7
Pairwise associations for <italic>Finger Pinching</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=499 hands).
[PNG File , 460 KB - jmir_v23i11e30394_app7.png]

Multimedia Appendix 8
Model selection for <italic>Finger Pinching</italic>: comparison of 4 mixed models of increasing complexity (parametric linear, quadratic and bounded growth models, and the nonparametric smoothing spline model) by root mean squared error and (effective) degrees of freedom used.
[PNG File , 329 KB - jmir_v23i11e30394_app8.png]

Multimedia Appendix 9
Pairwise associations for <italic>Draw a Shape</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=484 hands).
[PNG File , 325 KB - jmir_v23i11e30394_app9.png]

Multimedia Appendix 10
Linear quantile regression for <italic>Draw a Shape</italic> of short-term learning effects up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression <italic>P</italic> values are Bonferroni-adjusted (n=2420 tests).
[PNG File , 345 KB - jmir_v23i11e30394_app10.png]

Multimedia Appendix 11
Model selection for <italic>Draw a Shape</italic>: comparison of 4 mixed models of increasing complexity (parametric linear, quadratic and bounded growth models, and the nonparametric smoothing spline model) by root mean squared error and (effective) degrees of freedom used.
[PNG File , 325 KB - jmir_v23i11e30394_app11.png]

Multimedia Appendix 12
Pairwise associations for <italic>Two Minute Walk</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=171 patients).
[PNG File , 339 KB - jmir_v23i11e30394_app12.png]

Multimedia Appendix 13
Linear quantile regression for <italic>Two Minute Walk</italic> up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression <italic>P</italic> values are Bonferroni-adjusted (n=855 tests).
[PNG File , 303 KB - jmir_v23i11e30394_app13.png]

Multimedia Appendix 14
Pairwise associations for <italic>U-Turn</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=217 patients).
[PNG File , 346 KB - jmir_v23i11e30394_app14.png]

Multimedia Appendix 15
Linear quantile regression for <italic>U-Turn</italic> of short-term learning effects up to the fifth repetition. Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression <italic>P</italic> values are Bonferroni-adjusted (n=1085 tests).

[PNG File , 358 KB - jmir_v23i11e30394_app15.png]

Multimedia Appendix 16
Pairwise associations for <italic>Static Balance</italic> of the mean age, first score, fifth score, log-transformed number of repetitions performed, last score, and the difference from the fifth to the last score and their respective histograms (n=257 patients).
[PNG File , 298 KB - jmir_v23i11e30394_app16.png]

Multimedia Appendix 17
Linear quantile regression for <italic>Static Balance</italic> of short-term learning effects up to the fifth repetition (a smaller sway path is better). Comparison of baseline performance and linear slope of low (5th and 25th percentiles), median, and high performers (75th and 95th percentiles). Quantile regression <italic>P</italic> values are Bonferroni-adjusted (n=1285 tests).
[PNG File , 345 KB - jmir_v23i11e30394_app17.png]

Multimedia Appendix 18
Key results for the main analysis versus sensitivity analyses 1-3 for cognition, dexterity, and mobility.
[PDF File (Adobe PDF File), 754 KB - jmir_v23i11e30394_app18.pdf]

References

1. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, et al. Atlas of Multiple Sclerosis 2013: a growing global problem with widespread inequity. Neurology 2014 Sep 09;83(11):1022-1024 [FREE Full text] [doi: 10.1212/WNL.0000000000000768] [Medline: 25200713]
2. Kingwell E, Marriott JJ, Jetté N, Pringsheim T, Makhani N, Morrow SA, et al. Incidence and prevalence of multiple sclerosis in europe: a systematic review. BMC Neurol 2013 Sep 26;13:128 [FREE Full text] [doi: 10.1186/1471-2377-13-128] [Medline: 24070256]
3. Rao SM, Cognitive Function Study Group of the National Multiple Sclerosis Society. A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis. Medical College of Wisconsin, Milwaukee, WI. 1990. URL: https://scholar.google.com/scholar_lookup?title=A%20manual%20for%20the%20brief%20repeatable%20battery%20of%20neuropsychological%20tests%20in%20multiple%20sclerosis&publication_year=1990 [accessed 2020-10-07]
4. Parmenter BA, Weinstock-Guttman B, Garg N, Munschauer F, Benedict RH. Screening for cognitive impairment in multiple sclerosis using the symbol digit modalities test. Mult Scler 2007 Jan;13(1):52-57. [doi: 10.1177/1352458506070750] [Medline: 17294611]
5. Goodkin D, Hertsgaard D, Seminary J. Upper extremity function in multiple sclerosis: improving assessment sensitivity with box-and-block and nine-hole peg tests. Arch Phys Med Rehabil 1988 Oct;69(10):850-854. [doi: 10.5555/URI:PII:0003999388900093] [Medline: 3178453]
6. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, Multiple Sclerosis Outcome Assessments Consortium. The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):711-720 [FREE Full text] [doi: 10.1177/1352458517690824] [Medline: 28206826]
7. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, Multiple Sclerosis Outcome Assessments Consortium. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):704-710 [FREE Full text] [doi: 10.1177/1352458517690823] [Medline: 28206828]
8. D'Souza M, Papadopoulou A, Girardey C, Kappos L. Standardization and digitization of clinical data in multiple sclerosis. Nat Rev Neurol 2021 Feb;17(2):119-125. [doi: 10.1038/s41582-020-00448-7] [Medline: 33452493]
9. Montalban X, Graves J, Midaglia L, Mulero P, Julian L, Baker M, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. Mult Scler 2021 Jul 14:13524585211028561 [FREE Full text] [doi: 10.1177/13524585211028561] [Medline: 34259588]
10. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
11. Tortelli R, Simillion C, Lipsmeier F, Kilchenmann T, Rodrigues F, Byrne L, et al. The Digital-HD study: smartphone-based remote testing to assess cognitive and motor symptoms in Huntington's disease (1816). In: Proceedings of the American Academy of Neurology Annual Meeting. 2020 Apr 14 Presented at: American Academy of Neurology Annual Meeting; April 25-May 1, 2020; Online URL: https://n.neurology.org/content/94/15_Supplement/1816
12. Alexander S, Peryer G, Gray E, Barkhof F, Chataway J. Wearable technologies to measure clinical outcomes in multiple sclerosis: a scoping review. Mult Scler 2021 Oct;27(11):1643-1656 [FREE Full text] [doi: 10.1177/1352458520946005] [Medline: 32749928]

13. Lam K, Meijer K, Loonstra F, Coerver E, Twose J, Redeman E, et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. Mult Scler 2021 Aug;27(9):1421-1431 [FREE Full text] [doi: 10.1177/1352458520968797] [Medline: 33150823]
14. Bove R, White C, Giovannoni G, Glanz B, Golubchikov V, Hujol J, et al. Evaluating more naturalistic outcome measures: a 1-year smartphone study in multiple sclerosis. Neurol Neuroimmunol Neuroinflamm 2015 Oct 15;2(6):e162 [FREE Full text] [doi: 10.1212/NXI.0000000000000162] [Medline: 26516627]
15. Bradshaw MJ, Farrow S, Motl RW, Chitnis T. Wearable biosensors to monitor disability in multiple sclerosis. Neurol Clin Pract 2017 Aug;7(4):354-362 [FREE Full text] [doi: 10.1212/CPJ.0000000000000382] [Medline: 29185551]
16. Akhbardeh A, Arjona JK, Krysko KM, Nourbakhsh B, Gourraud PA, Graves JS. Novel MS vital sign: multi-sensor captures upper and lower limb dysfunction. Ann Clin Transl Neurol 2020 Mar;7(3):288-295 [FREE Full text] [doi: 10.1002/acn3.50988] [Medline: 32101388]
17. Midaglia L, Mulero P, Montalban X, Graves J, Hauser SL, Julian L, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: nonrandomized interventional feasibility study. J Med Internet Res 2019 Aug 30;21(8):e14863 [FREE Full text] [doi: 10.2196/14863] [Medline: 31471961]
18. Koh C, Lu W, Chen H, Hsueh I, Hsieh J, Hsieh C. Test-retest reliability and practice effect of the oral-format symbol digit modalities test in patients with stroke. Arch Clin Neuropsychol 2011 Jun;26(4):356-363. [doi: 10.1093/arclin/acr029] [Medline: 21555787]
19. Erlanger DM, Kaushik T, Caruso LS, Benedict RH, Foley F, Wilken J, et al. Reliability of a cognitive endpoint for use in a multiple sclerosis pharmaceutical trial. J Neurol Sci 2014 May 15;340(1-2):123-129 [FREE Full text] [doi: 10.1016/j.jns.2014.03.009] [Medline: 24656433]
20. Benedict RH, Cohan S, Lynch SG, Riester K, Wang P, Castro-Borrero W, et al. Improved cognitive outcomes in patients with relapsing-remitting multiple sclerosis treated with daclizumab beta: results from the DECIDE study. Mult Scler 2018 May;24(6):795-804 [FREE Full text] [doi: 10.1177/1352458517707345] [Medline: 28485186]
21. Cohen JA, Cutter GR, Fischer JS, Goodman AD, Heidenreich FR, Jak AJ, et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch Neurol 2001 Jun;58(6):961-967. [doi: 10.1001/archneur.58.6.961] [Medline: 11405811]
22. van Beek J, Freitas R, Bernasconi C, Montalban X, Butzkueven H, Kappos L, et al. FLOODLIGHT Open - A global, prospective, open-access study to better understand multiple sclerosis using smartphone technology. In: Proceedings of the Annual Meeting of the Consortium of Multiple Sclerosis Centers (CMSC). 2019 Presented at: Annual Meeting of the Consortium of Multiple Sclerosis Centers (CMSC); May 28–June 1, 2019; Seattle, Washington, USA URL: https://cmsc.confex.com/cmsc/2019/mediafile/Handout/Paper5900/QOL10_CMSC%202019%20FLOODLIGHT%20Open%20Poster_Global_van%20Beek%20et%20alFINA.pdf
23. Schwab P, Karlen W. A deep learning approach to diagnosing multiple sclerosis from smartphone data. IEEE J Biomed Health Inform 2021 Apr;25(4):1284-1291. [doi: 10.1109/JBHI.2020.3021143] [Medline: 32877343]
24. Solari A, Radice D, Manneschi L, Motti L, Montanari E. The multiple sclerosis functional composite: different practice effects in the three test components. J Neurol Sci 2005 Jan 15;228(1):71-74. [doi: 10.1016/j.jns.2004.09.033] [Medline: 15607213]
25. Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc 2004 Sep;99(467):673-686. [doi: 10.1198/016214504000000980]
26. Wood SN. Generalized Additive Models: An Introduction with R. Boca Raton: Chapman and Hall/CRC; 2017.
27. timwoelfle / MS-Practice-Effects. GitHub. URL: https://github.com/timwoelfle/MS-Practice-Effects [accessed 2021-11-02]
28. Complete dataset. Floodlight Open. URL: https://dataset.floodlightopen.com/public-blobs-prod/complete_dataset.csv [accessed 2021-11-02]
29. Liao S, Moor CD, Williams J, Fisher E, Rao S. Real-world practice effects and effect modifiers of neuroperformance tests in multiple sclerosis partners advancing technology and health solutions (MS PATHS) (4135). In: Proceedings of the American Academy of Neurology Annual Meeting. 2021 Presented at: American Academy of Neurology Annual Meeting; 2021; Online URL: https://n.neurology.org/content/96/15_Supplement/4135
30. Morrow S, O'Connor P, Polman C, Goodman A, Kappos L, Lublin F, et al. Evaluation of the symbol digit modalities test (SDMT) and MS neuropsychological screening questionnaire (MSNQ) in natalizumab-treated MS patients over 48 weeks. Mult Scler 2010 Nov;16(11):1385-1392. [doi: 10.1177/1352458510378021] [Medline: 20739335]
31. Benedict RH, Duquin JA, Jurgensen S, Rudick RA, Feitcher J, Munschauer FE, et al. Repeated assessment of neuropsychological deficits in multiple sclerosis using the symbol digit modalities test and the MS neuropsychological screening questionnaire. Mult Scler 2008 Aug;14(7):940-946. [doi: 10.1177/1352458508090923] [Medline: 18573822]
32. Roar M, Illes Z, Sejbaek T. Practice effect in symbol digit modalities test in multiple sclerosis patients treated with natalizumab. Mult Scler Relat Disord 2016 Nov;10:116-122. [doi: 10.1016/j.msard.2016.09.009] [Medline: 27919477]
33. Benedict RH, Smerbeck A, Parikh R, Rodgers J, Cadavid D, Erlanger D. Reliability and equivalence of alternate forms for the symbol digit modalities test: implications for multiple sclerosis clinical trials. Mult Scler 2012 Sep;18(9):1320-1325. [doi: 10.1177/1352458511435717] [Medline: 22277740]

34. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, Multiple Sclerosis Outcome Assessments Consortium. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):721-733 [FREE Full text] [doi: 10.1177/1352458517690821] [Medline: 28206827]
35. Cohen JA, Fischer JS, Bolibrush DM, Jak AJ, Kniker JE, Mertz LA, et al. Intrarater and interrater reliability of the MS functional composite outcome measure. Neurology 2000 Feb 22;54(4):802-806. [doi: 10.1212/wnl.54.4.802] [Medline: 10690966]
36. Maillart E, Labauge P, Cohen M, Maarouf A, Vukusic S, Donzé C, et al. MSCopilot, a new multiple sclerosis self-assessment digital solution: results of a comparative study versus standard tests. Eur J Neurol 2020 Mar;27(3):429-436. [doi: 10.1111/ene.14091] [Medline: 31538396]
37. Pelletier D, Rennert T, Chang M, Decunto S, Cen S. myMS: a comprehensive patient-centered mobile app to monitor MS at home (P32-003). In: Proceedings of the American Academy of Neurology Annual Meeting. 2019 Jul 05 Presented at: American Academy of Neurology Annual Meeting; 2019; Pennsylvania, Philadelphia, USA URL: https://n.neurology.org/content/92/15_Supplement/P3.2-003
38. Guo C, Cahir-Mcfarland E, Dumail J, Guymard T, Mcloughlin J, Bieuvelet S, et al. Konectom™ digital smartphone based assessment of cognitive and motor functional performance in multiple sclerosis. In: Proceedings of the MS Virtual 2020 Conference. 2021 Apr Presented at: MS Virtual 2020 Conference; December 1, 2020; Online p. S121. [doi: 10.1016/j.neurol.2021.02.360]
39. Naegelin Y, Lorscheider J, Woelfle T, Pless S, Kuhle J, Reyes O, et al. Dreams - Developing a comprehensive, sensitive and validated set of digital biomarkers for MS. In: Proceedings of the MS Virtual 2020 Conference. 2020 Presented at: MS Virtual 2020 Conference; December 1, 2020; Online. [doi: 10.1177/1352458520974937]

Abbreviations

9HPT: 9-hole peg test
ANOVA: analysis of variance
***e-SDMT*:** *electronic Symbol Digit Modalities Test*
MS: multiple sclerosis
RMSE: root mean squared error
SDMT: Symbol Digit Modalities Test
T25FW: timed 25-foot walk

Edited by R Kukafka; submitted 12.05.21; peer-reviewed by V Block, M Lindemann; comments to author 28.06.21; revised version received 23.08.21; accepted 12.09.21; published 18.11.21.

Please cite as:
Woelfle T, Pless S, Wiencierz A, Kappos L, Naegelin Y, Lorscheider J
Practice Effects of Mobile Tests of Cognition, Dexterity, and Mobility on Patients With Multiple Sclerosis: Data Analysis of a Smartphone-Based Observational Study
J Med Internet Res 2021;23(11):e30394
URL: https://www.jmir.org/2021/11/e30394
doi:10.2196/30394
PMID:34792480

Review

Conceptualizations of Cyberchondria and Relations to the Anxiety Spectrum: Systematic Review and Meta-analysis

Sandra K Schenkel[1], Dipl-Psych; Stefanie M Jungmann[1], Jun-Prof, Dr; Maria Gropalis[1], Dr; Michael Witthöft[1], Professor, Dr

Department of Clinical Psychology, Psychotherapy, and Experimental Psychopathology, Johannes Gutenberg-University Mainz, Mainz, Germany

Corresponding Author:
Sandra K Schenkel, Dipl-Psych
Department of Clinical Psychology, Psychotherapy, and Experimental Psychopathology
Johannes Gutenberg-University Mainz
Wallstraße 3
Mainz, 55122
Germany
Phone: 49 61313939202
Email: s.schenkel@uni-mainz.de

Abstract

Background: *Cyberchondria* describes the detrimental effects of health-related internet use. Current conceptualizations agree that cyberchondria is associated with anxiety-related pathologies and may best be conceptualized as a safety behavior; however, little is known about its exact underlying mechanisms.

Objective: This systematic review and meta-analysis aims to give an overview of the conceptualizations of cyberchondria and its relation to anxiety-related pathologies, quantify the strength of association to health anxiety by using meta-analyses, highlight gaps in the literature, and outline a hypothetical integrative cognitive-behavioral model of cyberchondria based on the available empirical evidence.

Methods: A systematic literature search was conducted using PubMed, Web of Science, and PsycINFO electronic databases. A total of 25 studies were included for qualitative synthesis and 7 studies, comprising 3069 individuals, were included for quantitative synthesis. The meta-analysis revealed a strong association of cyberchondria (r=0.63) and its subfacets (r=0.24-0.66) with health anxiety.

Results: The results indicate that cyberchondria is a distinct construct related to health anxiety, obsessive-compulsive symptoms, intolerance of uncertainty, and anxiety sensitivity. Further studies should distinguish between state and trait markers of anxiety-related pathologies and use experimental and naturalistic longitudinal designs to differentiate among risk factors, triggers, and consequences related to cyberchondria.

Conclusions: Health-related internet use in the context of health anxiety is best conceptualized as health-related safety behavior maintained through intermittent reinforcement. Here, we present a corresponding integrative cognitive-behavioral model.

(J Med Internet Res 2021;23(11):e27835) doi:10.2196/27835

KEYWORDS

cyberchondria; health anxiety; online health information seeking; anxiety; systematic review; meta-analysis

Introduction

Background

The internet allows anonymous access to a huge amount of specific information and opinions from nearly everywhere, at any time, and at relatively low costs [1,2]. It is increasingly being used to research health-related questions. Approximately 60% to 80% of internet users search the web for health-related information [3-5]. Of all internet search queries, 2% have medical content [6]. In 2013, 35% of surveyed American adults reported that they had started at least one web-based search session with the specific purpose of figuring out which medical condition they themselves or another person suffered from [7]. Arguably, a web-based search for perceived symptoms or feared illnesses can be helpful. Lemire et al [8] found that individuals feel empowered (ie, competent and in control) by conducting health-related internet use. In addition, health-related internet use seems to enable patients to take a more active role in patient-doctor–relationships [9]. It opens new opportunities for

illness prevention and health care, such as developing statistical models based on individual queries that might be used as a warning system, for example, for cancer [10]. However, there is also evidence of detrimental effects, such as an increase in worries or (health) anxiety, and as a consequence, an increase in the use of health care resources [11,12] (eg, 46% of a general population sample stated that health information found on the web led them to think they needed an appointment with a medical professional [7]).

In this context, journalists have coined the term *cyberchondria* [4,13] from the words *cyber*, referring to internet use, and *hypochondriasis*, referring to pathological health anxiety (HA). It is the belief in or the fear of having a serious disease, often without a matching medical condition. However, the term *cyberchondria* itself does not say anything about causality, the nature of the relationship between these two constructs, or the relevance of cyberchondria for patients with pathological HA. The most common definition of cyberchondria provided by Starcevic and Berle [14] postulates a bidirectional relationship. Elevated HA triggers health-related internet use, which in turn leads to amplification and, in the long term, to the maintenance of HA. Others classify cyberchondria-related behavior as a form of reassurance seeking that initially leads to an immediate decrease of HA but maintains HA in the long term through negative reinforcement [15].

Associations between cyberchondria-specific behavior and anxiety-related pathologies other than HA have also been proposed. First, a connection to obsessive-compulsive symptoms has been hypothesized [16-19], such that individuals experiencing greater cyberchondria also experience greater obsessive-compulsive symptoms. The interruption of other activities because of health-related internet use appears to be common in both constructs. Second, uncertainty about the seriousness of bodily symptoms and their appraisal as dangerous were proposed to be triggers for health-related internet use [1,14]. Several studies found medium-sized positive correlations between cyberchondria and intolerance of uncertainty (Multimedia Appendix 1 [20-23]). Furthermore, intolerance of uncertainty has been shown to moderate the impact of health-related internet use on HA [20,24]. The results suggest that the desire to avoid uncertainty and negative reactions to uncertainty are strongly associated with the experience of negative affective states because of health-related internet use [21,25]. Positive correlations were also found between anxiety sensitivity (ie, the tendency to interpret anxiety-related symptoms as signs of impending danger) and cyberchondria (Multimedia Appendix 2 [21,22,25]). Furthermore, anxiety sensitivity predicts cyberchondria in addition to the contributions of intolerance of uncertainty [25].

Different conceptualizations and theories have been developed that focus on certain aspects, such as emotional or behavioral consequences, cognitions, or characteristics of the search process itself; however, no consensus has yet been reached. The fact that cyberchondria is a current topic is also made clear by the fact that in 2019, for example, a systematic narrative review [26] and a meta-analysis of cyberchondria and HA [27] were published. Brown et al [26] presented an integrative cognitive-behavioral therapy (CBT) model of health-related internet use that distinguishes between reassuring health-related internet use on the one hand and problematic as well as compulsive health-related internet use on the other hand, according to the emotional consequences of this behavior. McMullan et al [27] found a meta correlation between HA and cyberchondria of r=0.62 (P<.001; n=10 studies).

Objective

This review aims to (1) give an overview of the current state of research regarding existing theoretical conceptualizations of cyberchondria and its relation to anxiety-related pathologies (ie, broader compared with the previous focus on HA), (2) quantify the strength of associations between cyberchondria and HA by using a meta-analysis according to the current data situation (ie, year 2020), (3) highlight gaps in the current literature, and (4) outline a hypothetical integrative cognitive-behavioral model of cyberchondria based on the available empirical evidence. This model follows a new approach by integrating existing results in the context of elevated HA and by waiving the artificial conceptual separation of health-related internet use according to the valence of its emotional effects.

Methods

Protocol and Search Strategy

A systematic literature search was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [28,29], using the computerized databases PubMed, Web of Science, and PsycINFO. They were searched three times, first in February 2016, second in July 2017, and third in February 2020, to include the latest research findings. A keyword search was performed using the following search terms and logics: *cyberchondria, cyberchondriasis, health-related internet use AND health anxiety, health-related internet use AND hypochondriasis, illness-related internet use AND health anxiety,* and *illness-related internet use AND hypochondriasis*, as well as the German translations of these terms. Searches were restricted to study titles and abstracts.

Inclusion and Exclusion Criteria

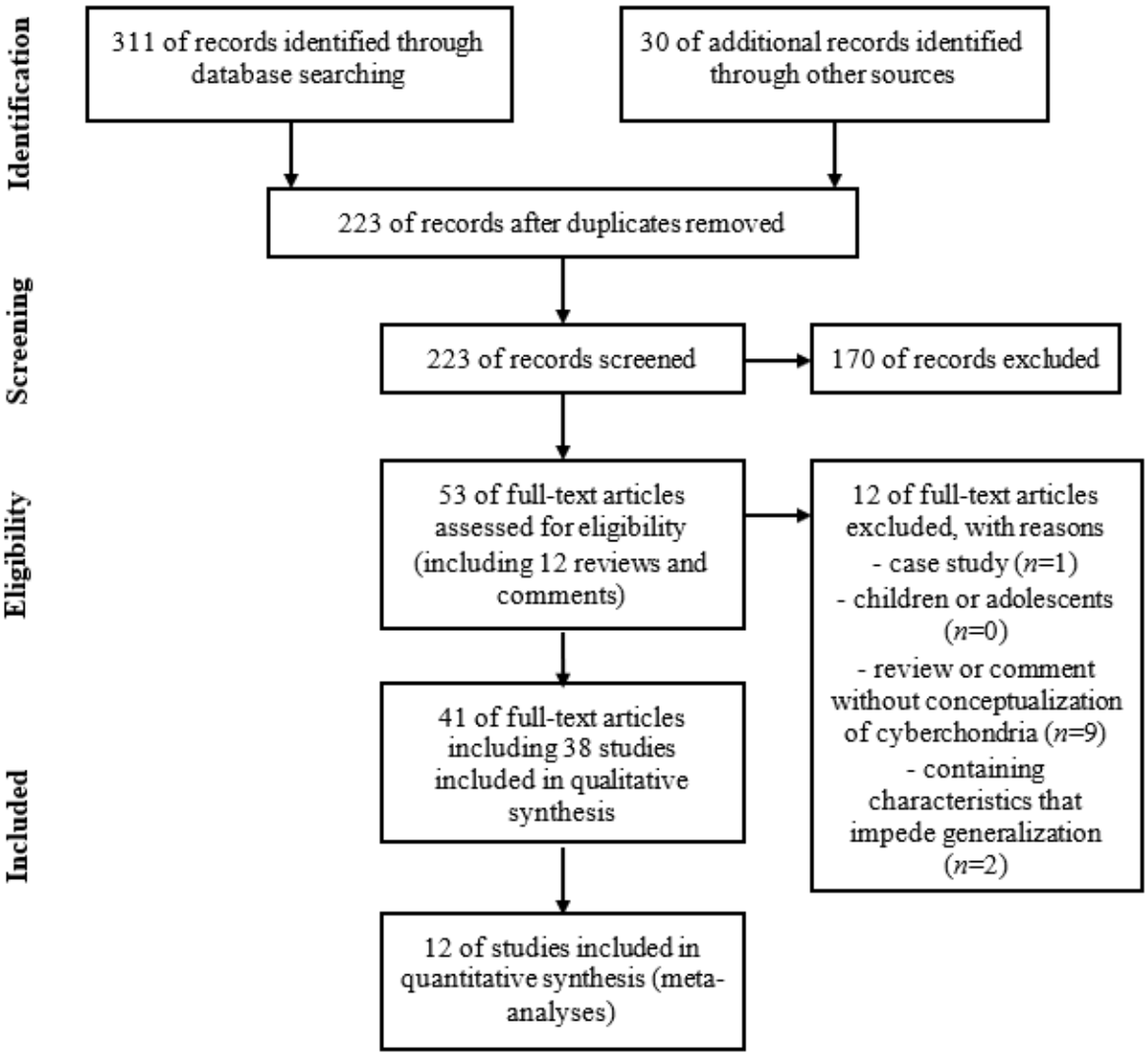
Only a few eligibility and exclusion criteria were specified to make this search as inclusive as possible. Studies were included if they (1) examined health-related internet use in the context of HA, (2) were published research papers or accepted manuscripts, (3) were written in English or German, and (4) investigated a general population or a health-anxious sample. Articles were excluded if they (1) reported a case study or data exclusively from children or adolescents or both, (2) were literature reviews or comments that did not postulate their own conceptualization or definition of cyberchondria, or (3) contained special characteristics that impeded the generalization of results (eg, investigation of a sample with symptoms outside the anxiety-related spectrum). All included research articles were already published in peer-reviewed journals.

Data Extraction and Synthesis Process

First, study titles were screened. Then, the abstract, method, and discussion sections were read consecutively (especially to

detect assumptions concerning the concept or function of cyberchondria), followed by reading full texts. Studies that did not fulfill all eligibility criteria or that fulfilled one or more exclusion criteria were eliminated. The electronic search was supplemented by reading the reference lists of the retrieved studies to identify more potential literature. Then, information from each eligible study was extracted and tabulated. Extracted data included names of the authors, publication year, journal, country of origin, type of study (primary vs secondary), proposed conceptualization of cyberchondria (yes or no), research question, study design, number of reported studies per paper, sample characteristics (age, range, type, and total sample size), methodology (operationalization of the construct, type of hypothesis, and analysis), and main findings. No paper was excluded because of insufficient information being reported. Applying these criteria to the searches resulted in the inclusion of 38 studies (Figure 1). All studies proposing a conceptualization of cyberchondria were then picked (irrespective of whether the study was primary or secondary). Different elements of conceptualization (eg, suggestion of detrimental effects of health-related internet use) were then extracted, which, in turn, were used as categories to group the corresponding primary studies (eg, studies investigating the effects of health-related internet use). Accordingly, one primary study could be assigned to more than one group. All primary studies were assigned. In the following sections, we present the studies according to the assigned groups (ie, *Conceptualizations of Cyberchondria*, *Cyberchondria and Anxiety-Related Pathologies*, *Triggers for Health-Related Internet Use*, and *Consequences of Health-Related Internet Use*).

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



Results

Conceptualizations of Cyberchondria

Conceptualizations of cyberchondria differ in the proposed immediate impact of health-related internet use on HA, and thus, also regarding the function of cyberchondria-related behavior in the context of HA (ie, *cyberchondria as an Amplifier of HA* vs *Cyberchondria as a Safety Behavior*). Multimedia Appendix 3 [6,14,15,26,30,31] summarizes the current conceptualizations.

Cyberchondria as an Amplifier of HA

Cyberchondria as a Failed Safety Behavior

Starcevic and Berle [14] presented a definition of cyberchondria as "an excessive or repeated search for health-related information on the internet, driven by distress or anxiety about health, which only amplifies such distress or anxiety"; moreover, "it does not denote a diagnosis and occurs as part of health anxiety and hypochondriasis." While giving a theoretical and empirically based framework, this definition also reformed the operationalization of cyberchondria by defining a process that includes a causal bidirectional relationship between cyberchondria and HA. It is important to underline that defining the precursors (HA as a trigger) and consequences (exacerbation of HA) of health-related internet use differentiates cyberchondria from merely searching the web for health-related information. Moreover, the multidimensionality of the construct was implied (eg, distress, excessiveness, and compulsion) [17]. Starcevic and Berle [14] hypothesized that cyberchondria-related behavior may be a form of reassurance seeking but that it should not be classified as a *classic* safety behavior, as in the short term, it increases anxiety instead of decreasing it [32]. If this definition is correct, it is important to explain how health-related internet use increases HA and why this behavior does not stop despite its adverse effects. In this regard, factors that amplify anxiety during health-related internet use are presented [33], which are assumed to set vicious circles in motion. They include the misinterpretation of the rank order of search results as pointing to the likely cause of a searched symptom [6,34] and the questionable trustworthiness of web-based health information [35]. However, current and partly inhomogeneous results raise doubts in this context [13,34,36].

Query Escalation

White and Horvitz [6] hypothesized that a process referred to as *query escalation* is highly relevant for increased HA during medical web-based searches. They defined cyberchondria as "the unfounded escalation of concerns about common symptomatology, based on the review of search results and literature on the Web." Escalations, in turn, were defined as special cases of cyberchondria during health-related internet use. They were operationalized as a shift of attention concerning the content of the information sought: away from a probable cause (eg, dehydration) for a common symptom (eg, a headache) to a serious illness (eg, cancer), which is a very unlikely cause for the bodily sensation.

To investigate this process, 11,158 naturalistic logs of a large (N=515) nonclinical sample's internet searches were analyzed [6]. Approximately 2% of all queries had medical content, and 5.3% of them escalated, thus verifying the existence of query escalations. However, as the direct effects on HA were not assessed, it could not be shown that this was a mediating process for experiencing increased anxiety. Evidence was found that searches were performed repeatedly, as 78.3% of all medical queries contained a symptom that was sought again within 2 weeks after the initial search. Moreover, the results indicated that health-related internet use occurred impulsively with a switch between longer abstinent episodes and episodes of intensive searching behavior, which might be indicative of an intermittent reinforcement pattern.

Singh and Brown [37] investigated the occurrence and consequences of query escalations in dependence on the HA level of their participants in a laboratory experiment. They found that escalations resulted in higher anxiety compared with nonescalations, regardless of previously existing HA, and highly health-anxious individuals were significantly more likely to escalate. Approximately 60% of all participants performed at least one escalated health-related internet use.

There are initial indications for the relevance of *query escalation* as a moderating mechanism for the negative effects of health-related internet use and these lead to the conclusion that attentional aspects might be of relevance in cyberchondria. In a consecutive study, White and Horvitz [34] showed that potentially alarming content (eg, *heart attack* and *medical emergency*) in captions, snippets, and URLs of search results influences the search-result click-through behavior so that this attribute makes them more likely to be selected.

Dimensional Conceptualizations of Cyberchondria

Cyberchondria was usually assessed using single items referring to the frequency or duration as well as the emotional consequences of health-related internet use [4,13,20]. These operationalizations neither took into account the postulated process character nor the multidimensionality of this construct [6,14]. In 2014, McElroy and Shevlin [17] developed the Cyberchondria Severity Scale (CSS), designed as a continuous measure of distress because of health-related internet use (according to previous conceptualizations [4,6,13,14]). This self-assessment questionnaire comprises 33 items that ask questions on a 5-point Likert scale about the frequency of web-based health-related behaviors. In contrast to previous assessment strategies, the CSS consists of five interpretable factors that reflect different dimensions of cyberchondria [16,17,38,39]. The factor *compulsion* reflects the different ways in which a web-based health-related search can unintentionally interrupt other web-based and offline activities. The dimension *distress* mirrors subjective negative emotional states because of health searches on the internet (stress, worry, anxiety, panic, and irritation). The factor *excessiveness* captures repeated searches for the same topic and the use of numerous sources. These 3 dimensions were also proposed by Starcevic and Berle [14]. Two additional subfacets are presented: all items loading on the factor *reassurance seeking* (hereafter called *reassurance*) indicate the felt need for reassurance from medical professionals triggered by information found on the web, representing a behavioral consequence of cyberchondria. The factor *mistrust of medical professionals* (hereafter called *mistrust*) reflects the conflict that web searchers have concerning whether to trust their medical professional over their own research results and self-diagnosis.

As detailed in Multimedia Appendix 4 [16-19,26,40,41] all CSS subscales showed moderate to strong intercorrelations, except for the *mistrust* factor, where only small or nonexistent associations were observed.

Although factor analyses revealed a five-factor structure of the CSS items, further analyses and observed correlation patterns raised doubts concerning the affiliation of *mistrust* to the concept of cyberchondria [16,38,39]. Higher-order factor analyses showed that a model consisting of a separate *mistrust* factor and a bifactor model comprising a general cyberchondria factor and 4 independent dimensions (*compulsion*, *distress*, *excessiveness*, and *reassurance*) best fit the data [16,38]. This showed that models generating a general cyberchondria factor had a superior fit to the data when *mistrust* was excluded. The four-factor structure was also found in a network analysis [42] that additionally revealed that no symptom seemed to be more central to the cyberchondria construct than others. On the other hand, Norr et al [38] pointed out that the findings regarding the *mistrust* factor may be attributable to method variance as all items of the scale are designed inversely. However, this approach has not been further investigated. Furthermore, qualitative data suggest that a certain mistrust of medical professionals and, specifically, of doctors, may play an important role [1]. In this context, a distinction has to be made between the affiliation of *mistrust* to the CSS and the affiliation of mistrust of doctors to cyberchondria; however, this question cannot be answered at this time. Further investigation is needed that takes into account the possibility that the CSS may not assess the construct in its entirety and that the general cyberchondria factor may represent a dimension of cyberchondria that is only weakly associated with *mistrust*. The original and revised CSS total scores (removing the *mistrust* items) were highly correlated (r=0.99 [16]).

Regarding convergent validity, the most important indicator for the CSS is its potential association with HA (see section *Health Anxiety*). Correlation analyses demonstrated that the CSS data were more strongly associated with *anxiety* (r=0.43; P≤.01) than with *stress* (r=0.37; P≤.01) and *depression* (r=0.24; P≤.01 [17]) using the short version of the Depression, Anxiety and Stress Scale [43]. The degree of association varies according to the theoretical similarities of the different constructs. Moreover, Fergus [16] showed that the CSS total score and the total score revised were more strongly associated with HA (r=0.59 and r=0.58, respectively; both P<.01) than with obsessive-compulsive symptoms (for both total scores, r=0.49; P<.01), as assessed using the Dimensional Obsessive-Compulsive Scale (DOCS) [44]. Barke et al [39] found medium positive correlations between the CSS global score and somatic symptoms (r=0.40; P<.01; assessed with the somatic symptom scale of the Patient Health Questionnaire [45,46]), depressive symptoms (r=0.31; P<.01; assessed with the short form of the Center for Epidemiologic Studies Depression Scale [47,48]), and health care use (r=0.29; P<.01; assessed with the Health Care Utilization questionnaire [49]).

As can be seen in Multimedia Appendix 4, the CSS subscales show different internal consistencies ranging from a high level for *compulsion* and *distress* to a good level for *excessiveness*, *reassurance*, and *mistrust*. The CSS total scale showed excellent internal consistency and split-half reliability without exception (α=.93-.96 [16,17,19,38,39]).

The first results show that the CSS seems to be unrelated to age [39], whereas results regarding sex differences are heterogeneous [2,39]. Too little information is currently available to finally evaluate these factors.

Other versions of the CSS regarding language (German [39] and Polish [40]) and length (short versions in English [50] and German [39]) show psychometric qualities comparable with the original version.

Jokić-Begić et al [51] developed a short questionnaire (4 items) to assess cyberchondria as a safety behavior. Besides predictors and negative outcomes, positive consequences were also intended to be measured; however, analyses revealed an unclear factor structure.

Cyberchondria as a Form of Problematic Internet Use

Problematic internet use describes the excessive use of the internet [52] for purposes other than searching for medical information and an inability to control that use [53], therefore reflecting behaviors from the fields of compulsion and addiction. Fergus and Dolan [30] define cyberchondria as a form of problematic internet use that is intended to reduce negative emotions but actually leads to greater subjective distress. To investigate this assumption, they used the Compulsive Internet Use Scale [54] and operationalized cyberchondria as the impact of health-related internet use on HA assessed with a single item. Individuals experiencing increased HA because of health-related internet use reported significantly greater levels of problematic internet use (mean 25.35, SD 11.58) compared with individuals experiencing no impact on (mean 18.63, SD 10.56; Cohen d=0.61; P<.01) or a decrease in HA (mean 21.76, SD 10.43; Cohen d=0.33; P<.01). These group differences stayed robust even after controlling for the frequency of health-related internet use and negative affect. These findings were supported by Fergus and Spada [31], who found a strong association between cyberchondria (operationalized by the CSS) and problematic internet use (r=0.59; P<.001). In this study, problematic internet use was assessed using the Problematic Internet Use Questionnaire [55], which consists of 3 dimensions (obsession, neglect of other activities, and control disorder). In addition, multiple linear regression with N=337 participants showed the robustness of this relationship (step 2; β=.41; P<.001) by controlling for age, gender, physical health, negative affect, and HA, where HA was the only covariate that accounted for variance in cyberchondria scores. Besides, Starcevic et al [42] found a stronger relationship between cyberchondria and problematic internet use than with HA. However, their network analysis showed that cyberchondria and problematic internet use were related but distinct constructs.

Fergus and Spada [31] concluded that cyberchondria could be viewed as a specific form of problematic internet use, indicating that concepts of problematic internet use may be of relevance to understanding cyberchondria. In addition, cognitive-behavioral treatments for problematic internet use may also be used to treat cyberchondria [30].

Moreover, Singh and Brown [11] found significant positive correlations between HA and 6 indicators of addiction to health-related internet use (inter alia, unsuccessfully trying to cut back, negative feelings regarding a real or anticipated loss, negative consequences, and increasing use over time).

Together, these findings underline the relevance of the compulsive and addictive aspects of cyberchondria-specific behavior.

Metacognitive Conceptualization of Cyberchondria

Fergus and Spada [22,31] hypothesized the conceptualization of cyberchondria according to the metacognitive model of emotional disorders [56]. Following this model, individuals engage in self-regulation strategies that maintain and worsen negative affective states because of their metacognitive beliefs. In terms of cyberchondria, this would mean that health-related images, memories, or thoughts trigger HA. At the same time, metacognitive beliefs are activated, which can be distinguished into two types: *positive* metacognitions associated with advantages of health-related worries (eg, "Considering all possibilities will help keep my mind at rest" [31]) and *negative* metacognitive beliefs associated with disadvantages or the uncontrollability of health-related worries (eg, "My thoughts are uncontrollable" [31]). To reduce the triggered HA, a self-regulation process, namely health-related internet use, is initiated. Plans for supposed successful self-regulation are represented as mentioned in metacognitive beliefs and additionally in beliefs about rituals and stop signals. Rituals refer to plans for coping with aversive inner emotional experiences, and stop signals refer to self-relevant goals that signal when to stop the self-regulation process. Negative metacognitive beliefs, especially regarding the performance of health-related internet use (eg, "Once I start searching I cannot stop" [31]), which can arise in the course, may lead to an increase in negative affect and, thus, lead to further self-regulation processes in the form of health-related internet uses, resulting in a repetitive and distress-evoking vicious circle.

The relevance of metacognitive aspects for cyberchondria was supported by several results from 3 web-based questionnaire surveys by the same authors [22,31], in which the Metacognitions Questionnaire–Health Anxiety (MCQ-HA; [57]) was used. It incorporates three metacognitive beliefs: thoughts can cause illness (MCQ-HA-C), biased thinking (MCQ-HA-B; eg, "Worrying about my health will help me cope"), and thoughts are uncontrollable (MCQ-HA-U). In all studies, zero-order correlation analyses revealed that all metacognition subscales were significantly correlated with cyberchondria (MCQ-HA-C: *r*s=0.32-0.49, all *P*<.001; MCQ-HA-B: *r*s=0.47-0.58, all *P*<.001; and MCQ-HA-U: *r*s=0.51-0.66, all *P*<.001).

Hierarchical multiple linear regression analyses statistically controlling for relevant covariates (ie, age, gender, physical health status, HA negative affect, and anxiety sensitivity) showed that *negative* metacognitions (ie, uncontrollability of thoughts [22,31], beliefs about rituals [22], and beliefs about stop signals [22]) and obsessive-compulsive symptoms [22] contributed significantly to the variance in cyberchondria. However, the analyses yielded inhomogeneous findings regarding the relevance of *positive* metacognitive beliefs [22,31]. Interestingly, neither anxiety sensitivity nor intolerance of uncertainty shared significant, unique associations with cyberchondria [22]. A supplemental analysis using HA as the criterion variable (instead of cyberchondria) and cyberchondria as a covariate revealed that beliefs about rituals and stop signals distinguish cyberchondria from HA.

In summary, the presented results support the importance of metacognitive aspects in the conceptualization of cyberchondria. Nevertheless, the proposed function as a predisposing and maintaining factor in the cyberchondria process needs to be further investigated. In particular, the heterogeneous findings regarding the relevance of *negative* (vs *positive*) metacognitive beliefs need further attention to clarify if they have a unique function in maintaining health-related internet use despite its adverse effects.

Cyberchondria as a Safety Behavior

The cognitive-behavioral model of HA and hypochondriasis by Warwick and Salkovskis [58] postulates that the perception of normal bodily symptoms and their misinterpretation as harmful and as a sign of a serious illness as well as the mechanism of somatosensory amplification [59] lead to negative affect (eg, anxiety and worry). Safety behavior is exhibited to reduce the negative affective states. Bleichhardt and Weck [15] classified health-related internet use as a new form of reassurance-seeking behavior that is carried out on the internet rather than by reading books or magazines or visiting a physician.

Support for this categorization of cyberchondria was given by Newby and McElroy [60]. They adapted regular strategies for reducing safety behavior to cyberchondria and integrated them into an internet-delivered CBT for HA. Strategies comprised, for instance, increased awareness of the frequency and personal cost of health-related internet use, activity scheduling, and reduction of excessive health-related internet uses through a behavioral experiment. Analysis of variance analyses revealed a significant reduction in cyberchondria pre- and post-CBT (within-group Hedges g=1.57, 95% CI 1.05-2.09). This reduction was greater than that in an active control group (Hedges g=1.1, 95% CI 0.60-1.61; time by group, $F_{1,67}$=25.41; *P*<.001). Mediation analyses showed that decreases in HA were partly mediated by a reduction in cyberchondria. Nevertheless, these promising results have to be interpreted in the light of several limitations, such as the influence of other therapeutic strategies and the moderate correlative association between cyberchondria and HA. Changes in cyberchondria might be reflective of changes in HA rather than of a unique process.

In 2019, the first CBT model was presented by Brown et al [26] based on a comprehensive review of existing data accounting for reassurance-related and compulsive elements.

According to the model, health-related internet use is commonly performed to eliminate health threats and is moderated by metacognitive beliefs (eg, "I need HIU to control my anxiety"), among other factors. Depending on the results of the health-related internet use, two possible forms of problematic health-related internet use can occur: first, if the search results can eliminate the perceived threat, relief is experienced, and health-related internet use is negatively reinforced. Consequently, health-related internet use is terminated, and metacognitive beliefs regarding the beneficial effects of this strategy are developed. This form is termed *pathological health-related internet use* and describes health-related internet

use functioning as a reassurance-seeking behavior in the context of HA (ie, cyberchondria, although the authors recommend not to use this term because of ambiguities in the definition). A second possible outcome occurs if search results strengthen the perceived threat. Resulting anxiety and worry lead to the continuation of health-related internet use, even to the point of query escalation. In consequence, metacognitive beliefs about the internet search itself and its possible negative consequences (eg, "I can't control my HIU") are developed and lead to distress. This form is termed *compulsive health-related internet use*. The perceived threat does not focus on health (in contrast to pathological health-related internet use) but on the internet search itself. Individuals affected then might feel stuck and out of control regarding health-related internet use.

Cyberchondria and Anxiety-Related Pathologies

In this section, show the associations between several anxiety-related pathologies (ie, HA, obsessive-compulsive symptoms, intolerance of uncertainty, anxiety sensitivity, and pain catastrophizing) and health-related internet use. Moreover, we discuss the distinctiveness of cyberchondria to anxiety-related pathologies.

Health Anxiety

All the aforementioned concepts and theories contain interrelations between HA and cyberchondria. For instance, Starcevic and Berle [14], as well as Brown et al [26], postulated that individuals who previously suffered from elevated HA are especially likely to experience negative affective states because of health-related internet use. Therefore, the following section illuminates the results regarding the relationship between HA and cyberchondria and is subdivided by the existing operationalizations of cyberchondria (by the CSS, including our meta-analyses, and by frequency and duration). As HA is seen as a dimensional rather than a categorical construct [61,62], findings from its whole spectrum are relevant.

A Meta-analytic Integration of Cyberchondria Operationalized by the CSS

As cyberchondria is supposed to be a multidimensional construct, it is conceivable that its different dimensions have dissimilarly strong relations to HA [38]. Multimedia Appendix 5 [18,19,38-40,63,64] displays the corresponding correlation coefficients. Individual dimensions have differing but significant associations with HA, ranging from high correlations with *distress* and *excessiveness*, moderate to high correlations with *reassurance* and *compulsion*, to moderate correlations with *mistrust*. To quantify the associations between (1) HA as a trait and cyberchondria, as well as between (2) HA and the different dimensions of cyberchondria as operationalized by the CSS, we conducted several meta-analyses. This examination is of particular importance to test the validity of the cyberchondria concept, as postulated by Starcevic and Berle [14].

Description of Included Studies

For the described purpose, we examined studies included in this review that reported results regarding a correlational hypothesis between a standardized measure of HA and the CSS total scale and subscales. On the basis of the aforementioned results regarding the questionable affiliation of the *mistrust* factor with the cyberchondria construct (or at least with the CSS total scale), only results that did not include the 3 *mistrust* items were incorporated. We contacted the authors of the primary studies to obtain missing data (eg, coefficients regarding the CSS total scale or subscales and internal consistencies). They were included if available. Quality ratings were conducted for the included articles, whereby the quality criteria according to Brown and Reuber [65] were adjusted based on the characteristics of the available studies (eg, some criteria were also met by all studies included in this study, such as correlative design and standardized measuring instruments, which were then not used as criteria). To assess the quality of the studies and especially the generalizability of the results, the following criteria were used and evaluated (by the 2 authors SKS and SMJ): consecutive sampling; validity items or data check, for example, missing values; control of possibly confounding variables; sufficiently reported inclusion and exclusion criteria; sufficiently reported sample characteristics; and approximate representativeness of the sample. According to Brown and Reuber [65], these ratings were used to calculate the overall quality of the study methods, which was defined as the proportion of items given a *yes* rating in combination with a sufficiently large sample size. We rated the sample size according to Cohen [66], who suggested that 85 participants were needed to detect a medium effect size (r=0.3), given an α level of .05 and power of 0.8. The overall quality was rated as high (≥80% *yes*, equating to no more than one of the methodological standards given a *no* rating and N≥85), medium (50-79% *yes* and N≥85), low (20-49% *yes* and N≥85), and unacceptable (<20% *yes* or N<85). Any unacceptable studies were excluded from the meta-analyses. Interrater reliability (between the 2 authors SKS and SMJ) across all categories was κ=0.75, indicating a substantial interrater agreement [67]. Quality ratings are provided in Multimedia Appendix 6 [16,21,22,25,31,39,40,63,64]. Of the 12 studies rated, 8 (67%) were judged to be of high quality, 3 (25%) of medium quality, 1 (8%) of low quality, and no study was rated to be of unacceptable quality, resulting in 12 studies that were included in the first meta-analysis regarding the first question (Table 1). Owing to unavailable missing data, 17% (2/12) of these studies were not considered in the meta-analysis regarding the second question, resulting in 10 studies for the second question.

Table 1. Details of studies included in the meta-analyses to quantify the association between HA[a] and cyberchondria operationalized by the CSS[b].

First author, year	Effect size, *r*	N	Age, mean (SD)	Sex (female), n (%)	HA (Cronbach α)	HA total score, mean (SD)	CSS (Cronbach α)	Ex[c]: medical condition	In[d]: performing health-related internet use	Country
Fergus, 2014 [16]	0.59	539	31.3 (9.9)	234 (43.4)	SHAI[e] (.91)	11.12 (6.91)	CSS (.96)	Yes	Yes	United States
Fergus, 2015 [21]	0.62	578	31.2 (9.8)	253 (43.7)	WI[f] (.91)	13.27 (5.06)	CSS (.95)	Yes	No	United States
Norr, Albanese et al, 2015 [25]	0.53	526	34.9 (12.4)	364 (69.2)	SHAI (.9)	14.40 (7.86)	CSS (.95)	No	No	United States
Barke et al, 2016, study A [39]	0.605	499	29.1 (10.4)	367 (73.6)	mSHAI[g] (.9)	—[h]	German CSS (.93)	No	No	Germany
Barke et al, 2016, study B [39]	0.581	292	24.2 (4.1)	223 (76.4)	mSHAI (.93)	—	German short form of the CSS (.83)	No	No	Germany
Fergus and Russell, 2016 [19]	0.51	375	31.6 (10.2)	177 (47.3)	MIHT[i] (.9)	—	CSS (.95)	Yes	No	United States
Fergus and Spada, 2017, study 2 [31]	0.67	260	32.9 (9.2)	106 (40.8)	WI (.92)	13.66 (5.48)	CSS (.95)	No	Yes	United States
Fergus and Spada, 2018, study 1 [22]	0.56	330	19.4 (2.1)	220 (66.6)	WI-6 (.89)	—	CSS (.95)	No	Yes	United States
Fergus and Spada, 2018, Study 2 [22]	0.61	331	38.7 (10.4)	177 (53.5)	WI-6 (.92)	—	CSS-15–Revised (.88)	No	Yes	United States
Bajcar et al, 2019[j] [40]	0.56	240	26.5 (11.1)	203 (57.1)	SHAI (.93)	—	CSS-PL (.95)	No	No	Poland
Gibler et al, 2019[j] [63]	0.58	221	19.2 (1.7)	156 (70.6)	SHAI (.89)	10.86 (6.12)	CSS (.95)	No	No	United States
Mathes et al, 2019 [64]	0.61	462	36.56 (12.9)	297 (64.3)	SHAI (.92)	TI[k]: 7.18 (4.41); FOI[l]: 3.13 (3.00)	CSS (.96)	No	No	United States

[a]HA: health anxiety.

[b]CSS: Cyberchondria Severity Scale.

[c]Ex: exclusion criterion.

[d]In: inclusion criterion.

[e]SHAI: Short Health Anxiety Inventory.

[f]WI: Whiteley Index.

[g]mSHAI: modified version of the Short Health Anxiety Inventory.

[h]Missing data.

[i]MIHT: multidimensional inventory of hypochondriacal traits.

[j]Excluded from the second meta-analysis (association between dimensions of cyberchondria and HA) because of unavailable data.

[k]TI: subscale thought intrusion of the SHAI.

[l]FOI: subscale fear of illness in SHAI.

Study Characteristics

Most studies were conducted in the United States (9/10, 90%). HA was mostly operationalized by the Short Health Anxiety Inventory (SHAI [68]) or a short version (7/10, 70%). Approximately 40% (4/10) used the Whiteley Index [41], and 10% (1/10) used the multidimensional inventory of hypochondriacal traits [69].

Participant Characteristics

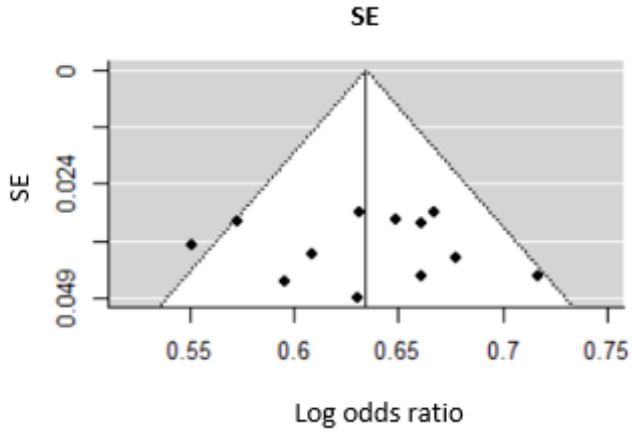
In total, 4653 participants were included in the meta-analyses, with a mean age of 30.5 years; 58.3% were women.

Approximately 30% (3/10) of studies excluded participants who reported that they suffered from a diagnosed medical condition, and 40% (4/10) of studies only included participants who stated that they regularly performed health-related internet use.

Data Analysis

The random-effects model, considered to be the most appropriate in applied sciences was chosen [70,71]. This model is based on the assumption that the population effect differs randomly from study to study and accounts for within- and between-study variability [72,73]; therefore, more general inferences can be made compared with fixed-effect models. For statistical evaluation, the method provided by Hunter and Schmidt [74,75] was applied using the *metafor* software package in R (R Foundation) [76], following Viechtbauer's [77] recommendations. This method aims to investigate relations on the level of constructs rather than measured values; therefore, it allows correction for the effects of several statistical artifacts, inter alia, concerning measuring accuracy (or more specifically for internal consistencies [78]). Potential publication bias was explored using a funnel plot (Figure 2). The Egger regression asymmetry test was used to detect publication bias. These analyses were also performed using the *metafor* software package in R [76]. For all the results, a two-sided *P* value of ≤0.05 was considered significant.

Figure 2. Funnel plot of integrated studies.



Meta-analytic Results

First, correlation coefficients were corrected for measurement error (Cronbach α [74]) of the instruments used to assess HA and cyberchondria. To quantify the association between these two constructs (question 1), a model with no moderators was estimated, revealing a small heterogeneity of integrated effects (I^2=27.8%; Q_{11}=16.7; *P*=.12 [79]); so no moderator analysis was conducted. A strong positive association between HA and cyberchondria was found (*r*=0.63; *P*<.001, 80% CI 0.61-0.66). Figure 3 shows the integrated effects. The Egger regression analysis showed that publication bias was not present (z=0.25; *P*=.80). Another meta-analysis concerning the strong positive association between HA and cyberchondria closely corresponded to our findings (*r*=0.62; *P*<.001 [27]).

To quantify the association between the different dimensions of cyberchondria and HA (question 2), a second model was estimated that integrated the corrected correlation coefficients between HA and all subscales of the CSS (Figure 4).

In line with our expectations, high heterogeneity of effects was found (I^2=95.59%; Q_{44}=1020.81; *P*<.001). Therefore, subscales were included in the model as moderators and accounted for R^2=67.26% of the variance. Table 2 shows the results of this meta-regression analysis. All subscales showed significant, small to strong associations with HA, and most strongly with *distress* and *excessiveness*.

Figure 3. Forest plot of integrated study effects and meta-effect regarding the association between cyberchondria and health anxiety.

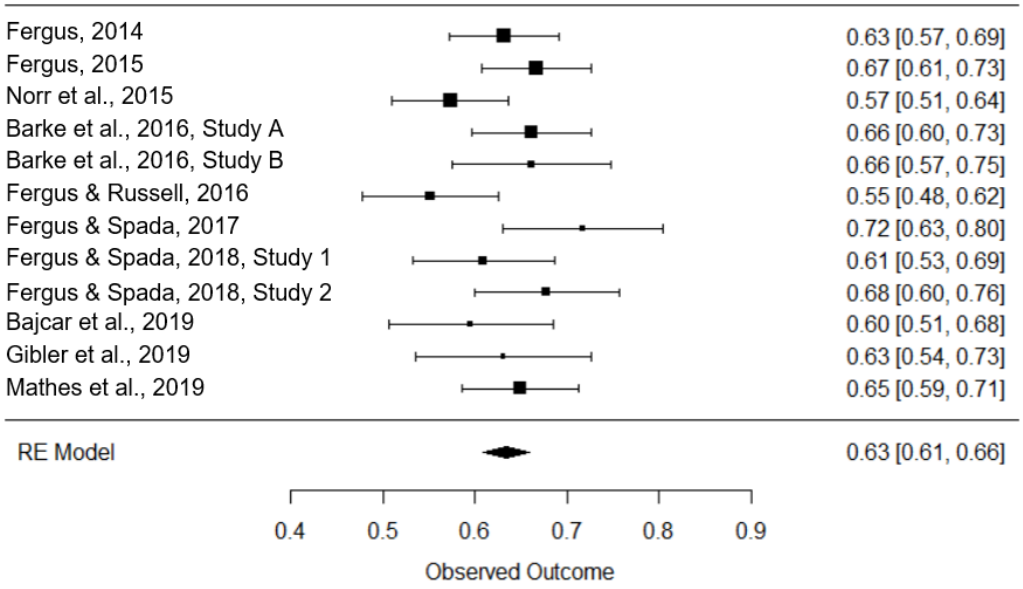


Figure 4. Forest plot of integrated study effects regarding the association between cyberchondria dimensions and health anxiety.

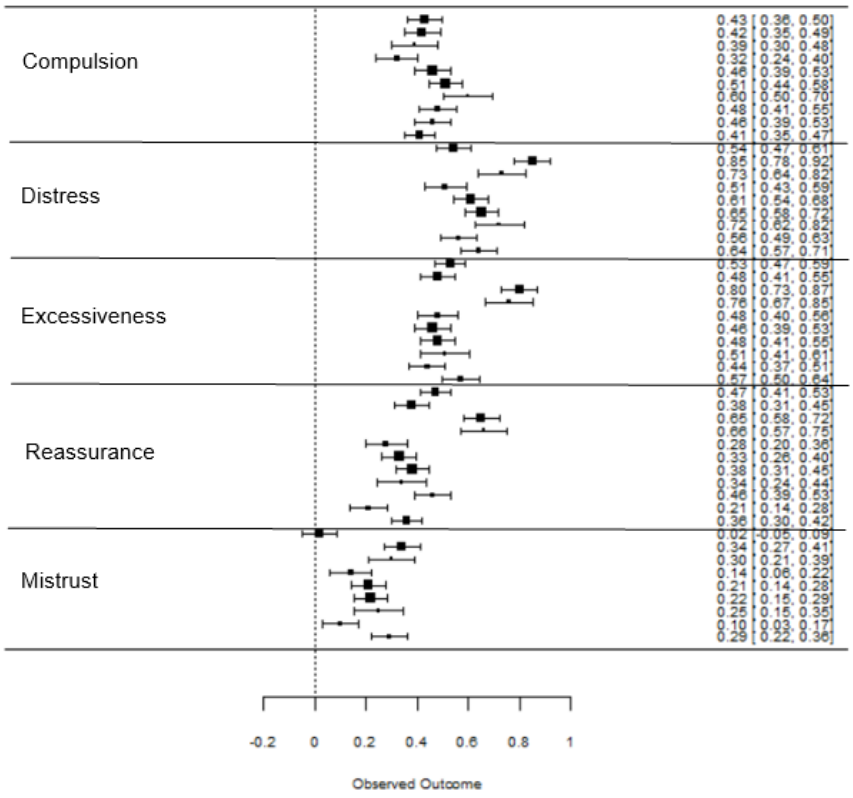


Table 2. Results of meta-regression to quantify the association between cyberchondria dimensions and health anxiety.

Dimension of cyberchondria	β	*P* value	95% CI
Compulsion	.46	<.001	0.23-0.69
Distress	.66	<.001	0.43-0.89
Excessiveness	.59	<.001	0.36-0.83
Reassurance	.44	<.001	0.21-0.67
Mistrust	.24	<.05	0.01-0.47

Correlational Results Regarding Subfacets of HA

Correlation analyses further revealed that cyberchondria clusters most strongly correlated with the affective component of HA, more precisely with health worries (r=0.57; P<.01 [19]; Multimedia Appendix 5). This finding is consistent with the results of Norr et al [38], who found relations between cyberchondria and the SHAI, which is argued to primarily assess an affective component of HA [24]. As worry was found to be linked to safety behavior [80], the authors suggest that health-related internet use may be an activity performed to reduce health worries.

Cyberchondria Operationalized by Frequency and Duration of Health-Related Internet Use

Until the publication of the CSS [17], cyberchondria was assessed in two parts using two items: (1) frequency or duration of health-related internet use, and (2) its effects. Corresponding studies are detailed in Multimedia Appendix 7 [4,11,13,20,36,81,82]. Mostly moderate positive correlations were reported, indicating that the more health-anxious a person was, the more frequently and longer he or she searched the web for health- and illness-related information. This association was also found when HA was not operationalized continuously but dichotomously by ranking the SHAI scores of participants and using the bottom and top quartiles [4]. TePoel et al [81] observed a moderate positive association between HA and health-related internet use, indicating that individuals with higher HA reported more health-related internet use. This was true for both HA groups, clinical (β=.26; P<.001) and nonclinical (β=.29; P<.001).

Obsessive-Compulsive Symptoms

It is hypothesized that cyberchondria is connected with obsessive-compulsive symptoms. Correlational studies have used DOCS [44] to assess these. It consists of 4 dimensions: *contamination*, *responsibility for harm* (hereafter called *responsibility*), *unacceptable thoughts* (hereafter called *thoughts*), and *symmetry*. Each is judged for time spent, avoidance, distress, interference, and attempts to control. It was shown that the DOCS total scale is significantly correlated to the CSS total scales with (r=0.49, P<.01 [16]; r=0.38, P<.01 [40]) and without *mistrust* (r=0.49, P<.01 [16]; r=0.56, P<.001 [22]; r=0.38, P<.0 [40]) and to all CSS subscales (*distress:* r=0.50, P<.01 [16]; r=0.43, P<.01 [40]; *compulsion:* r=0.46, P<.01 [16]; r=0.34, P<.01 [40]; *excessiveness:* r=0.35, P<.01 [16]; r=0.31, P<.01 [40]; *reassurance*: r=0.27, P<.01 [16]; r=0.23, P<.01 [40]; *mistrust:* r=0.11, P<.01 [16]; r=0.07, not significant [40]), such that individuals with greater obsessive-compulsive symptoms experience greater cyberchondria. Moreover, as detailed in Multimedia Appendix 8 [18,19,40], all CSS and DOCS subscales showed significant and mostly moderate intercorrelations. No statement can be made about the *mistrust* factor as it was excluded from the primary analyses. Interestingly, correlation coefficients were consistently smaller in the study conducted by Fergus and Russell [16] compared with those by Norr et al [18]. This might originate from a range restriction because of stricter eligibility criteria used by Fergus and Russell [19] regarding the performance of health-related internet use and the nonexistence of a diagnosed medical condition.

Although an association between cyberchondria and obsessive-compulsive symptoms was confirmed, different hypotheses exist about the nature of this relationship. First, it is supposed that obsessive-compulsive symptoms are an essential part of the cyberchondria construct. As was seen in section *Dimensional Conceptualizations of Cyberchondria*, McElroy and Shevlin [17] argue that health-related internet use can interrupt other web-based and offline activities. The results regarding the CSS dimension *compulsion* are reported in the aforementioned section. Here, we recap the results briefly. The subscale correlates highly with the CSS total scale (r=0.82; P<.01) and shows moderate to strong correlations with the other CSS subscales (r=0.26-0.80, excluding *mistrust*). Factor analyses [16,38] showed that *compulsion* seems to be part of a higher-order cyberchondria factor.

Another group of authors hypothesized that obsessional and compulsive behavioral elements contribute to the maintenance of cyberchondria [26]. Starcevic and Berle [14] suggested that maintenance is because of the combination of (1) an *obsessional doubt* regarding the validity and sufficiency of information found on the web, and (2) a hope to find the one *perfect* explanation for the symptoms experienced, which may represent a compulsive element of the cyberchondria construct. Following this argument, the subject would assume that continuing to perform health-related internet use would increase the probability of finding the *ultimate* answer, and for this purpose, any anxiety triggered by health-related internet use is acceptable. However, no empirical findings exist to address this hypothesis. Norr et al [18] postulated maintenance of health-related internet use because of a positive bidirectional association between cyberchondria and symptoms related to the DOCS subscale *contamination*. If health-related internet use is carried out to obtain reassurance, this behavior may lead to an increased obsession with physical health, owing to the strong link between cyberchondria and HA. This, in turn, may lead to an increased urge to prevent contamination and illnesses, for example, by increased hand washing. This amplified obsession with health may lead to further health-related internet use. Moreover, it is conceivable that if individuals with cyberchondria believe that health-related internet use gives them the power to prevent future illnesses, there may be a connection to the DOCS subscale *responsibility*. The *symmetry* and *thoughts* subscales are postulated to be unrelated. The structural equation model confirmed the predicted unique relationships. This model, controlling for negative affect (Positive and Negative Affect Schedule [83]) and HA (SHAI [68]), provided a good fit to the data (χ^2=2954.63; P<.05; CFI=0.94; Tucker–Lewis index=0.93; root mean square error of approximation=0.05; 90% CI 0.05-0.06). Norr et al [18] concluded that a co-occurrence of both constructs and repeated health-related internet use might function as a safety behavior to reduce contamination concerns or responsibility for harm.

Another approach by Fergus and Russell [19] hypothesized that the existing findings regarding associations are because of a redundancy of the constructs. Health-related internet use may represent the behavioral component of both HA (eg, reassurance

seeking) and obsessive-compulsive symptoms (eg, neutralizing behavior). They conducted confirmatory factor analyses to examine whether intercorrelations among these 3 constructs were best represented by indicator loadings on the same or separate latent constructs. Analyses showed that cyberchondria seems to be a distinct construct from HA and obsessive-compulsive symptoms, although they are related as a correlated three-factor model provided the best fit to the data ($\chi^2_{(41)}$=149.3; comparative fit index=0.96; nonnormed fit index=0.95; standardized root mean square residual=0.06; root mean square error of approximation=0.08; 90% CI 0.69-0.99). In this model, a latent correlation of r=0.58 (P<.001) between cyberchondria and obsessive-compulsive symptoms was observed. This finding of distinctness of both constructs was confirmed by a network analysis by Starcevic et al [42].

In summary, the available empirical evidence is insufficient to finally characterize the nature of the relationship between cyberchondria and obsessive-compulsive symptoms. Moreover, it remains unclear whether obsessive-compulsive symptoms contribute to the maintenance of cyberchondria or vice versa. One clear shortcoming of research to date is that exclusively cross-sectional questionnaire-based surveys exist, which do not allow the investigation of processes and causality. Experimental studies and longitudinal data are needed to answer these questions.

Intolerance of Uncertainty

Intolerance of uncertainty is defined as a dispositional fear of the unknown [84]. It captures the inability to tolerate the uncertainty of ambiguous situations. It was found to comprise 2 dimensions, which are operationalized in the multidimensional Intolerance of Uncertainty Scale [85]: *prospective* intolerance of uncertainty reflects perceptions of threat as well as implications of future uncertainty. *Inhibitory* intolerance of uncertainty mirrors the desire to avoid uncertainty and behavioral symptoms of apprehension when faced with it. Considering intolerance of uncertainty as a risk factor for cyberchondria seems to be justified because of its substantial associations with HA [25,86].

The results from a qualitative interview study by Singh et al [1] support the relevance of uncertainty about whether one has an undiagnosed health issue for developing a felt need to go on the web and search for medical content. An association was confirmed by mainly medium-sized correlations (Multimedia Appendix 1), and a network analysis also revealed that both constructs were distinct [42]. In particular, *inhibitory* intolerance of uncertainty may be a risk factor for experiencing distress because of health-related internet use [21,25].

In addition, during health-related internet use, multiple medical possibilities are presented to the user [6], and for people with high intolerance of uncertainty, ambiguous situations were found to be distressing. Moreover, creating catastrophic interpretations of unclear health information is only related to HA at a high intolerance of uncertainty [24]. The moderating role of intolerance to uncertainty on the relationship between cyberchondria and HA seems conceivable to explain why some individuals experience worse HA as a consequence of health-related internet use, whereas others do not (see section *Consequences of Health-Related Internet Use*). A regression analysis by Fergus [20] confirmed this finding. The relationship between the frequency of health-related internet use and HA because of these searches was significant at high levels of intolerance of uncertainty (b=0.27; partial r=0.21; P<.01) but not at low (b=–0.06; partial r=–0.03, not significant).

This indicates that the desire to avoid uncertainty and negative reactions to it are strongly associated with experiencing negative affective states because of health-related internet use, potentially because of increased catastrophic thinking about the meaning of bodily symptoms, as concluded by Fergus [21]. Another conceivable explanation for this finding may be that the tendency to react negatively to uncertainty and the resulting desire to avoid it may lead to further health-related internet use to reduce uncertainty. However, the more health-related internet uses are conducted, the greater the possibility of finding ambiguous and inconsistent information, which has the potential to trigger the negative affect.

In addition, intolerance of uncertainty was also found to moderate another detrimental effect of health-related internet use concerning anxiety sensitivity (see section *Anxiety Sensitivity*). Only individuals with high intolerance of uncertainty experience elevated anxiety sensitivity as a consequence of viewing web-based medical content [23].

Anxiety Sensitivity

Although people with intolerance of uncertainty experience uncertainty in general as dangerous, people with high levels of anxiety sensitivity interpret anxiety-related symptoms as dangerous [23]. This was found to be a potential risk factor for HA [87], and therefore, a link between anxiety sensitivity and cyberchondria-specific behavior was hypothesized [21,25]. It is commonly operationalized by the Anxiety Sensitivity Index-3 [88], which consists of three lower dimensions reflecting different types of concerns: *cognitive* (ie, concerns about mental incapacitations), *physical* (ie, concerns about immediate complications), and *social* (ie, concerns about social rejection because of publicly observable symptoms of anxiety). Correlational analyses supported the relevance of anxiety sensitivity for the cyberchondria construct (Multimedia Appendix 2). The coefficients range from a moderate to a high level. Different conclusions can be drawn from this study. The results may indicate that health-related internet uses are performed on various concerns (*cognitive*, *physical*, and *social*). Moreover, anxiety sensitivity may be relevant, especially for those experiencing distress because of health-related internet use. The structural equation models that included anxiety sensitivity, intolerance of uncertainty, and HA (operationalized by the SHAI) as independent variables to predict cyberchondria (operationalized by the CSS total score, excluding *mistrust*) highlighted the important role of anxiety sensitivity as a predictor for cyberchondria in addition to the contributions of intolerance of uncertainty and HA [25]. However, besides the role of anxiety sensitivity as a risk factor, it was also hypothesized to be a detrimental consequence of health-related internet use. Corresponding results are reported in detail in the section *Consequences of Health-Related Internet Use*.

Pain Catastrophizing

In the context of chronic pain, the contribution of pain catastrophizing to cyberchondria has been investigated [63]. It has been shown that pain catastrophizing (ie, the tendency to ruminate and worry about pain [89]) significantly predicts cyberchondria and its facets (operationalized by the CSS), even when controlling for negative affect and HA. Individuals with chronic pain often experience anxiety and distress about the origin and consequences of their pain perceptions and search for answers on the web. Pain catastrophizing may amplify the negative affect and may contribute to the initiation of the cyberchondria-specific vicious circle, as postulated by Starcevic and Berle [14]. Further support is provided by logistic regression analysis that aims to predict engaging in health-related internet use by psychopathologies and somatic symptoms [5] and comprises data from 992 adults from the general population of whom, 751 (75.7%) reported engaging in health-related internet uses in the past 3 months. It was shown that conducting health-related internet uses was associated with HA, obsessive-compulsive symptoms, and intolerance of uncertainty (among other psychopathologies); however, only the severity of somatic symptoms independently predicted health-related internet use when each of these pathological domains was controlled for.

Triggers for Health-Related Internet Use

Another relevant aspect in these models are the triggers for health-related internet use. For instance, Starcevic and Berle [14] postulated that health-related internet use is triggered by elevated levels of HA or distress. However, till now, research has paid little attention to the precursors, motivations, and triggers for health-related internet use. To our knowledge, only 2 qualitative studies have addressed this question, both using semistructured interviews with participants with high levels of HA (ie, participants' SHAI scores exceeded a critical cutoff, indicating clinically significant HA (N=8, SHAI>16 points [90]; N=20, SHAI>18 points [1]). The studies yielded similar results, which are reported in the following section and integrated as far as possible with the superordinate themes identified by McManus et al [90] (*information is power* and *novelty of internet searching*). If no literature reference is given in parentheses, the 2 studies report the same results.

"*Information is power*" describes a participant's hope to be able to prevent future illness, the deterioration of an already existing one [1], or death because of an illness [90] by gathering information, such as appropriate health promotion strategies or remedies [1]. In this context, Singh and Brown [1] found a positive association between HA and the frequency of using the internet to obtain *wellness* information. This can be interpreted as an attempt to prevent the onset of a serious disease, resulting in a feeling of being more in control of one's health. The internet is also used for self-diagnoses to find a label for a health issue so that more specific information can be collected [90]. The possibility of failing to prevent an illness leads to the anticipation of dreadful consequences, which increases the perceived urgency to conduct health-related internet use [1]. Moreover, uncertainty about the probability of having a serious disease primed illness-related thoughts and worries leading to heightened anxiety, whereas being uncertain about nonserious issues led to curiosity [1]. In addition, health-related internet use is initiated by negative experiences and expectations concerning health care professionals that may lead to mistrust: inconvenience, too short appointment times, lack of reassurance, and past negative experiences. Participants also use the internet as a first filter for information to prepare for or justify a consultation with a physician.

The second superordinate theme was *novelty of internet searching* and included the perceived advantages of the internet as a medium of information compared with other methods. Advantages were a low threshold for use, the possibility to share others' experiences and, in consequence, feel comforted by the knowledge that one is not the only person who suffers from this health issue [90], as well as speed, convenience, the possibility of obtaining opinions and advice from other persons affected, the huge size of the source, and its ability to provide reassurance. Reported disadvantages included the following: the information obtained was too broad, confusing, and often conflicting, as well as the presence of false, noncredible, or irrelevant information [1].

It can be concluded that a combination of different factors seems to lead to the initiation of health-related internet use. Besides situational factors (states) such as uncertainty and anxiety, more permanent or predisposing factors such as prior beliefs (eg, mistrust of medical professionals and *information is power*), seem to be important [1] as the metacognitive conceptualization of cyberchondria proposes [31]. There is still a lack of research investigating predisposing factors and triggers. Studies with naturalistic designs appear to be especially promising when investigating situational factors.

Consequences of Health-Related Internet Use

Emotional Consequences of Health-Related Internet Use

Negative Emotional Consequences of Health-Related Internet Use

The impact of health-related internet use on HA has mainly been assessed using questionnaires. Multimedia Appendix 9 [4,11,13,20,36,82] gives an overview of the results. The studies differ in terms of (1) the operationalizations of HA, health-related internet use, and its emotional effects; (2) the predictor for emotional effects (HA vs frequency of health-related internet use); and (3) the type of hypothesis and analysis (correlational vs difference). The results indicate that individuals with higher HA are more likely to experience negative emotional consequences after health-related internet use, which is in accordance with the study by Starcevic and Berle [14]. This finding is also supported by the results of the meta-regression that we conducted, as described in the section *A Meta-analytic Integration of Cyberchondria Operationalized by the CSS*, regarding the CSS subscale *distress* and its association with HA (β=0.66; P<.001). In addition, about one-third of individuals in general population samples stated that they felt increased anxiety after health-related internet use (31.4% [30]; 38.5% [91]). However, it should be noted that no measure of HA was included in this study.

To our knowledge, 2 experimental studies were conducted in this regard and supported the results of the questionnaire studies. Participants in a study by Baumgartner and Hartmann [13] read a text about a fictitious bacterial disease that is accompanied by ordinary symptoms such as stomach and intestinal pain. In addition, the trustworthiness of the source of information was manipulated. The information about the fictitious illness indeed increased negative responses in health-anxious individuals but only if the information was judged as trustworthy (perceived relevance of information: b=0.70, *P*<.01; chance of being already infected: b=0.94, *P*<.01; chance of getting infected: b=1.02, *P*<.01; and worry about disease: b=0.99, *P*<.01). Singh and Brown [37] showed that query escalations reduced postsearch anxiety immediately in 17.0% of the cases in the high HA group, remained the same in 27.65% of cases, and increased in 55.3% of cases. Proportions were comparable for the low HA group (χ^2_2=0.491; *P*=.78). Interview studies revealed that uncertainty, fright, anxiety, worry, and nervousness [1,90] could be experienced as a consequence of health-related internet use. Moreover, significant positive associations (*r*=0.31, *P*<.01 [19]; *r*=0.34, *P*<.001 [22]; *r*=0.44, *P*<.01 [63]) between cyberchondria (operationalized by the CSS total score) and negative affect (operationalized by the Positive and Negative Affect Schedule) were found, indicating that the higher the level of cyberchondria, more negative the affect experienced is. It should be noted that no statements can be made about a causal relationship.

Norr et al [23] hypothesized that being exposed to medical web-based information that is congruent with catastrophic interpretations of anxiety symptoms (namely, as a threat to health) may contribute to the development and maintenance of elevated anxiety sensitivity. They compared undergraduate students (N=52) in two conditions. The experimental group was exposed to medical websites converted to PDF files that provided catastrophic interpretations of bodily symptoms. The control group viewed the PDF files generated from general health and wellness websites. Hierarchical regression analyses showed that individuals in the experimental group experienced significantly higher anxiety sensitivity after the manipulation than did the control group (R^2=0.81, *P*≤.001 and ΔR^2=0.4, *P*≤.01). This relationship was moderated by the individuals' levels of intolerance of uncertainty, such that viewing medical websites only affected anxiety sensitivity in participants with high levels of intolerance of uncertainty. Therefore, it was concluded that health-related internet use and the concomitant exposure to medical content might call users' attention to potentially alarming causes of bodily sensations and may thereby fuel both HA and anxiety sensitivity [21,25].

Positive Emotional Consequences of Health-Related Internet Use

Besides the reported results confirming the postulated negative emotional consequences of health-related internet use, it should be noted that positive emotional consequences were also found, such as relief and calm [1,11,90]. The process of searching itself can provide reassurance by giving a person the feeling of having done something instead of ruminating [90]. As mentioned above, Singh and Brown [37] found that in escalated health-related internet use, anxiety after search was reduced in 17.0% of cases, even in individuals with high levels of HA. Doherty-Torstrick et al [82] found that a considerable proportion of their sample (32.8% in the high HA group and 71.2% in the low HA group) experienced no impact or decreased anxiety during and after health-related internet use. Fergus and Dolan [30] stated that 40.7% of their sample experienced reduced HA after health-related internet use. White and Horvitz [91] reported that 76% of their participants felt reassured, and 50.3% experienced reduced anxiety.

Predictors of Emotional Consequences of Health-Related Internet Use

Some of the factors influencing whether positive or negative emotional consequences are more likely experienced have already been discussed above (ie, HA, intolerance of uncertainty, anxiety sensitivity, and obsessive-compulsive symptoms). However, certain characteristics of the searching process itself also seem to have an influence. Substantial associations between the frequency and the duration of searches and elevated HA postsearch were found (Multimedia Appendix 9). Individuals engaging more often or longer in this behavior are more likely to experience negative consequences. McManus et al [90] found that the longer individuals searched the web, the more likely they were to find alarming information that, in turn, would lead to heightened anxiety. Another important factor seems to be the type of information found: participants stated that negative emotions occurred during health-related internet use if unfamiliar health issues were researched or potentially alarming information was found [1,90]. Conversely, positive emotions were experienced if familiar issues were researched or nonserious causes for experienced symptoms were found. Postsearch negative emotions occurred when people perceived the information as inconclusive, conflicting, confusing, or serious. In this case, new searches were initiated. Postsearch positive emotions were the result of finding a nonserious answer, diagnosing oneself, finding a remedy, or clarifying uncertainties. In these cases, the search was terminated. The results regarding query escalations support these findings [6,37]. Furthermore, the type of web-based service used seems to have an impact: using forums that provide web-based contact with experts or the possibility of sharing with other affected persons seems to alleviate anxiety, whereas using diagnosis systems and video platforms seems to have the opposite effect [36].

Reciprocity Between Health-Related Internet Use and HA

TePoel et al [81] investigated whether HA influences health-related internet use and vice versa. They chose a -wave longitudinal study design with 2-month time lags between the waves. Health-related internet use was operationalized by frequency. Analyses were conducted separately for individuals with lower levels of HA (N=4,564, Dutch version SHAI score<18) and pathological levels of HA (N=751, Dutch version SHAI score≥18); this is of particular importance regarding the distinctiveness of cyberchondria from HA and the question of whether the mechanisms underlying cyberchondria are different in individuals with different levels of HA. A random intercept cross-lagged panel model was chosen for the data analysis. This multilevel approach is able to control for stable (trait-like) individual differences by distinguishing between within-level

and between-level variances. In this model, two random intercepts were included for HA and health-related internet use. In the pathological HA group, changes in an individual's HA score were not predicted by a change in health-related internet use 2 months earlier. Conversely, changes in an individual's health-related internet use score were not predicted by changes in HA 2 months earlier. Thus, no evidence for a reciprocal relationship between HA and health-related internet use was found in individuals with pathological HA. In contrast, evidence for a reciprocal association was found in the nonclinical group with lower HA, such that increases in an individual's HA were predicted by increases in health-related internet use and vice versa.

This indicates that the underlying mechanisms for the impact of health-related internet use on HA differ depending on the previously existing level of HA. In the pathological HA group, health-related internet use seemed to be a maintaining rather than an increasing factor. This suggests that findings from samples with normal levels of HA should only be transferred to clinical samples with care. This highlights the importance of investigating the whole range of HA (especially clinical samples) and its immediate effects. The question of causality remains unanswered in this context.

Effects of Health-Related Internet Use on Behavior: Reassurance Seeking and Health Care Use

It has been hypothesized that health-related internet use is able to trigger further health care use to receive reassurance, especially in individuals with elevated HA. This hypothesis is supported by several findings. Singh and Brown [11] found a medium correlation between HA and the likelihood of visiting a physician after the search (r=0.22; P<.01). Our meta-regression (see section *A Meta-analytic Integration of Cyberchondria Operationalized by the CSS*) regarding the association between the CSS subscale *reassurance* and HA yielded a strong correlation (β=.44; P<.001). Eastin and Guinsler [92] investigated the moderating role of HA in this process. HA was operationalized by the Health Anxiety Questionnaire [93], health-related internet use by its frequency, and health care use because of health-related internet use by the frequency of physician consultations and the number of different physician visits. Regression analysis confirmed a significant interaction, indicating that as HA increases, the relationship between the frequency of health-related internet use and health care use increases. Furthermore, individuals with high HA (Illness Attitude Scale≥47 [41]) considered dysfunctional behavior (eg, physician hopping and ordering nonprescribed medication on the web) to be more likely because of health-related internet use [36]. In terms of the Cyberchondria model proposed by Brown et al [26], these results might indicate that individuals with normal levels of HA are better able to correctly judge the seriousness of the threat originating from health-related internet use and decide whether a physician consultation is necessary. Interestingly, it seems as if the use of physical health care plays a more important role compared with mental health care use. Using structural equation modeling, the CSS subscale *reassurance* showed a stronger association with the use of physical health care (β=.70) than with mental health care (β=.50) over the past 60 days [64]. In addition, the subscale *excessiveness* was negatively associated with mental health care use (β=–.49*;* P<.001), indicating that individuals who repeatedly engage in health-related internet use report less use of mental health care.

Effects of Health-Related Internet Use: Quality of Life and Functional Impairment

In an extension of the *compulsion* subfacet of cyberchondria [17], which reflects different ways in which health-related internet use can interrupt other web-based and offline activities, it was hypothesized that cyberchondria might lead to significant impairment in psychosocial functioning.

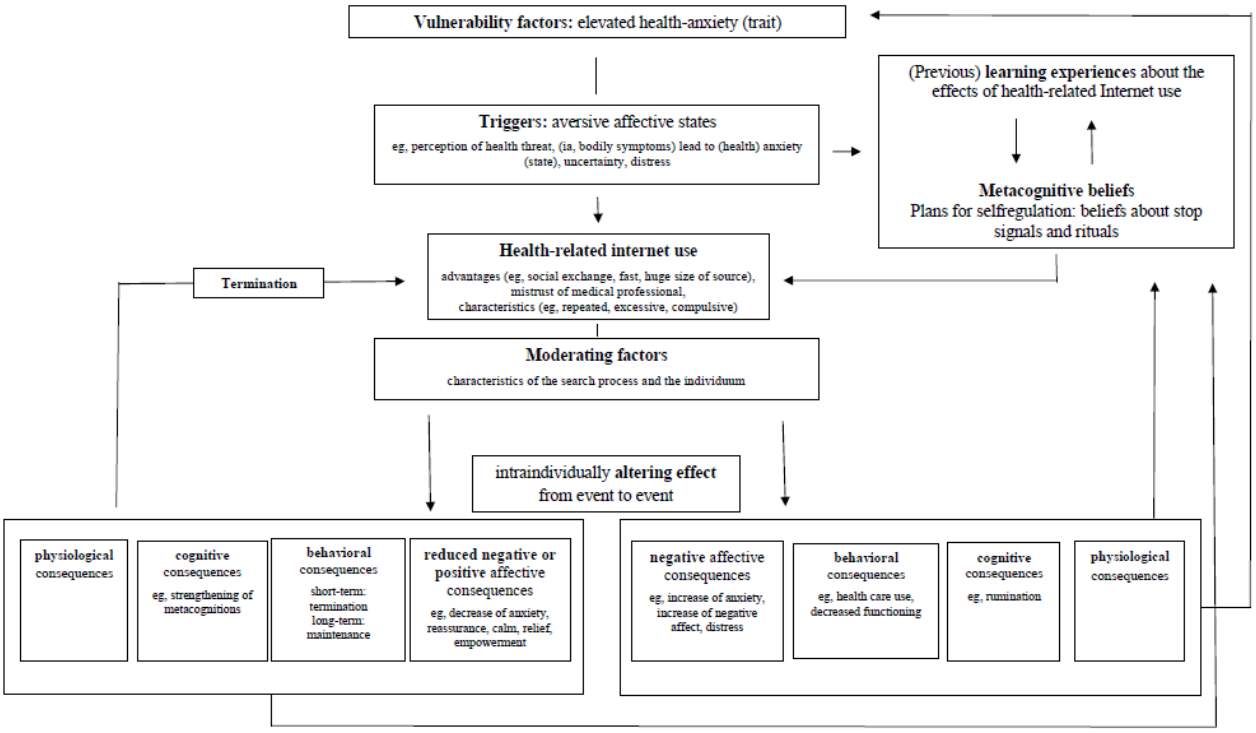
In line with this, it was found that cyberchondria was strongly associated with greater functional impairment even when controlling for the effects of HA (operationalized by the SHAI [68]) [64]. Functional impairment was defined as one's ability to engage in daily activities, operationalized using the Sheehan Disability Scale [94]. Interestingly, when accounting for HA, CSS was not associated with decreased quality of life, which was operationalized as one's overall level of contentment and satisfaction regarding 4 domains: physical and psychological health, social relationships, and environment (World Health Organization Quality of Life Assessment [95]). These results might indicate that individuals with cyberchondria may be satisfied with their lives, although they feel functionally impaired by their cyberchondria-specific behavior.

Discussion

Principal Findings

The aims of this work were to (1) give an overview of existing conceptualizations of cyberchondria and its relation to anxiety-related pathologies, (2) quantify its association with HA, and (3) highlight gaps in the current literature. In an attempt to summarize and reconcile the partly contradictory models and findings, an integrative hypothetical cognitive-behavioral model of cyberchondria as a health-related safety behavior that is maintained through intermittent reinforcement was developed (aim 4; Figure 5).

Figure 5. Hypothetical cognitive-behavioral model of cyberchondria as a health-related safety behavior maintained through intermittent reinforcement.



Previous conceptualizations of cyberchondria proposed that health-related internet use should be divided into *cyberchondria* and *classic safety behavior* based on its (emotional) impact. However, in the context of pathological HA, there is no evidence that health-related internet use followed by negative emotional consequences is conceptually different from health-related internet use followed by positive ones. This distinction marks the crucial point regarding the incompatibility of definitions and the integration of existing findings on emotional consequences. Regarding elevated levels of HA, we, therefore, propose the conceptualization of cyberchondria as a classic health-related safety behavior that is maintained by intermittent reinforcement. In addition, following Brown et al [26], we recommend not to use the term *cyberchondri*a any longer, inter alia, as this would imply a conceptual distinction between health-related internet use and other health-related safety behaviors. In the following, *cyberchondria* is used only to mark results that were assessed using previous conceptualizations (eg, the CSS). In the following, the components of our model are explicated in detail.

All existing theories suggest that cyberchondria-specific behavior is triggered by emotions that can be summarized as aversive emotional states. A more recent model [26] suggests that the perception of a health threat may be central to initiation. We also share this view in the field of elevated and pathological HA. Confronted with *situations* that are evaluated as threatening to health, highly health-anxious individuals are hypothesized to experience, for example, (health) anxiety, uncertainty, and distress. Such a situation may be, for instance, viewing illness-related materials or even the perception of bodily symptoms. As is known for pathological HA, there is also support for the view that this factor may be relevant. To mention, there are the substantial associations between cyberchondria and health concerns triggered by bodily sensations (as assessed with the Whiteley Index), anxiety sensitivity, and the severity of somatic problems (as seen in the context of chronic pain). However, to our knowledge, no study exists that examines the perception of bodily symptoms as a trigger.

Besides the emotional aspects, the perception of a health threat may have consequences on the cognitive (eg, health concerns) and physiological levels (ie, changes corresponding to the affective state). Regarding the latter point, the process of somatosensory amplification [59] may contribute to the initiation and maintenance of health-related internet use. At the behavioral level, health-related internet use may be conducted to reduce (concordant) negative affect. Interview studies give the first indications that health-related internet use is triggered by a combination of stable and situational factors, supporting the conceptualization of cyberchondria in a cognitive-behavioral model. To our knowledge, no studies have addressed the important question of immediate precursors (especially state HA) in an experimental or naturalistic study design.

The first existing questionnaire-based studies highlighted the relevance of metacognitions regarding cyberchondria-specific behavior. Moreover, the reported results of interview studies contained results that can be integrated into this context (eg, *Information is power*). In terms of content, metacognitions are usually distinguished as positive and negative with distinct functions (initiation vs maintenance of health-related internet use); however, the results are heterogeneous. Therefore, this separation was not included in our model. Following Fergus and Spada [22,31] as well as Brown et al [26], we also propose for the field of elevated and pathological HA that aversive negative affective states activate metacognitive beliefs that may represent plans for self-regulation strategies (namely health-related internet use), rituals, and stop signals that are

conducted to cope with aversive (affective) states. At this point, we expand existing theories by supplementing the interaction between metacognitions and previous learning experiences regarding the (short-term emotional) consequences of health-related internet use. The latter may include the experience that health-related internet use may have altering effects from event to event; that is, that besides negative emotions, positive consequences (eg, relief) can also occur. Therefore, on the one hand, metacognitions may develop from learning experiences (eg, "HIU is useful. At some point, I will feel better"). On the other hand, these metacognitions may influence how negative outcomes of health-related internet uses are processed and contribute to its maintenance.

Health-related internet use seems to have certain advantages compared with other health-related safety behaviors that may make it more likely for health-anxious individuals to select this strategy over others. Besides obvious aspects (ie, anonymity, huge amount of specific information, low costs, and promptness), social contact and exchange seem to be relevant. Regarding mistrust of professionals, partly contradictory results were reported, and the general recommendation of not to include this aspect in the construct of cyberchondria was critically discussed. Brown et al [26] included this factor as previous experiences in their model as a kind of vulnerability factor. We also see sufficient support for this but interpret mistrust in the context of why health-related internet use is selected over other safety behaviors. In addition, it is conceivable that this aspect may be more relevant in samples with pathological levels of HA, given that the relationships between physicians and highly health-anxious patients are often difficult and characterized by dissatisfaction [96].

In addition, numerous results exist regarding the characteristics of health-related internet uses once the vicious circle process is set into motion. The reported findings support the notion that health-related internet use occurs repeatedly and has excessive and compulsive subcomponents. Individuals with higher HA engage in health-related internet use for longer periods and more frequently (Multimedia Appendix 7). Moreover, they are more likely to repeat searches and use numerous sources, as well as to experience that health-related internet use interrupts other activities. Correlational and factor analysis examinations yielded further evidence (especially regarding the positive association between HA and the CSS subfacets, *excessiveness* and *compulsion*). Concerning compulsive elements, further supportive results were found: first, an association between HA and the 6 indicators of addiction to health-related internet use, and second, a strong association between problematic internet use (reflecting behaviors from the fields of compulsion and addiction) and cyberchondria.

Concerning the consequences of health-related internet use, all authors agree that immediately after health-related internet use, both positive emotional consequences (especially reduction of HA), as well as negative ones (increase of HA) can occur. Indeed, the existence of both valences was confirmed even in highly health-anxious individuals. Most previous studies have focused on interindividual differences concerning emotional consequences in cross-sectional study designs. With this approach, the effects of health-related internet use are indirectly assumed to be stable over time within an individual (almost trait-like); however, this has not yet been investigated. Future studies should follow an event-based approach to investigate possible changes in intraindividual effects over time. If individuals experience different effects of health-related internet use, altering from event to event, this may be an indicator of the presented conceptualization.

Previous findings suggest that individuals with higher degrees of anxiety-related pathologies (obsessive-compulsive symptoms, intolerance of uncertainty, and anxiety sensitivity) are more likely to experience cyberchondria, especially negative emotional consequences (Multimedia Appendix 9). Following TePoel et al [81], we hypothesize that an illness-related bias may be of certain relevance regarding the moderating role of emotional outcomes. It was found to occur in highly health-anxious individuals [97,98], as well as in patients with pathological HA [99-101] and describes the tendency to focus on information that confirms their health worries and to ignore information that is contradictory [58]. However, to our knowledge, no study has addressed this bias in the context of cyberchondria. In addition, the characteristics of the medical search session itself seem to influence emotional outcomes. The type of information found may also be relevant (alarming and unfamiliar and *query escalation*), as well as the greater frequency and longer duration that is more likely to produce elevated negative emotional states (Multimedia Appendix 9). The results regarding the moderating role of the perceived trustworthiness of a source are inhomogeneous and therefore not included in the model. We also propose the consequences of health-related internet use on other behavioral components corresponding to the affective outcome, as shown in Figure 5. As the results indicate that bodily symptoms may be a trigger, it seems conceivable that they increase and decrease corresponding to the emotional effects of health-related internet use. Cognitive and physiological consequences need to be further investigated.

We hypothesize that health-related internet use is maintained through intermittent reinforcement (ie, altering valences of effects). For the positive valence effects, we propose two possible ways of maintenance. First, through negative reinforcement, that is, in consequence of health-related internet use, aversive states may be reduced, such as the perceived health threat, concordant negative feelings (eg, HA), or aversive physiological (anxiety) symptoms. Second, through positive reinforcement, positive valence emotions may occur (eg, reassurance, relief, calm, empowerment, and increased sense of control). Regarding negative valence consequences, the above-described metacognitive beliefs ("Some time or other, I will feel better") may maintain health-related internet use besides its negative effects. In addition, it seems conceivable that amplification of aversive states may retrigger self-regulative behavior (ie, health-related internet use).

Strengths and Limitations

The strengths of this review include the updated systematic research strategy based on the PRISMA guidelines and the focus on the models underlying the studies. All potentially eligible studies were assessed for inclusion. The meta-analyses make

an important contribution to this field by aggregating the results of numerous studies and quantifying the association between cyberchondria and HA. The development of a hypothetical CBT model for elevated and pathological HA provides further starting points for future research and treatment. There were also limitations of the review. Although trying to make the literature search as inclusive as possible, the exclusion criteria may have resulted in the omission of relevant studies. Potential papers not written in English or German or those investigating children and adolescents were excluded, which may have resulted in linguistic, cultural, or age bias. Moreover, the generalizability of the meta-analytic results to the level of constructs is restricted: first, because of the inclusion of relatively few studies and the operationalization of cyberchondria and HA, and second, because samples were mainly recruited via the internet. This may lead to an artificially heightened homogeneity of samples, which restricts the search for moderators.

Conclusions

Cyberchondria-specific behaviors appear distinct from but strongly related to HA, intolerance of uncertainty, obsessive-compulsive symptoms, and anxiety sensitivity. Numerous findings support the dominant conceptualization of cyberchondria by Starcevic and Berle [14], which postulates a bidirectional relationship between health-related internet use and HA. However, no clear evidence exists for the key elements of previous definitions regarding the conceptual differentiation of cyberchondria as a classic safety behavior based on its emotional consequences. Future research needs to further investigate the immediate emotional consequences of health-related internet use, especially in individuals with pathological HA on an interindividual level, using experimental and naturalistic longitudinal study designs. Additional variables besides HA (eg, feelings of uncertainty) have to be taken into account when examining the mechanisms of initiation and maintenance that are mediated by the immediate effects of health-related internet use.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlations between intolerance of uncertainty and cyberchondria.
[DOCX File , 15 KB - jmir_v23i11e27835_app1.docx]

Multimedia Appendix 2
Correlations between anxiety sensitivity and cyberchondria.
[DOCX File , 14 KB - jmir_v23i11e27835_app2.docx]

Multimedia Appendix 3
Summary of currently existing conceptualizations of cyberchondria and their key elements.
[DOCX File , 15 KB - jmir_v23i11e27835_app3.docx]

Multimedia Appendix 4
Factor intercorrelations and internal consistencies of the Cyberchondria Severity Scale.
[DOCX File , 20 KB - jmir_v23i11e27835_app4.docx]

Multimedia Appendix 5
Correlations between health anxiety and the Cyberchondria Severity Scale subscales.
[DOCX File , 16 KB - jmir_v23i11e27835_app5.docx]

Multimedia Appendix 6
Quality ratings of integrated studies.
[DOCX File , 16 KB - jmir_v23i11e27835_app6.docx]

Multimedia Appendix 7
Association between health anxiety and cyberchondria operationalized by frequency and duration of health-related internet use.
[DOCX File , 19 KB - jmir_v23i11e27835_app7.docx]

Multimedia Appendix 8
Correlations between obsessive-compulsive symptoms and the Cyberchondria Severity Scale subscales.
[DOCX File , 14 KB - jmir_v23i11e27835_app8.docx]

Multimedia Appendix 9
Emotional effects of health-related internet use.
[DOCX File , 22 KB - jmir_v23i11e27835_app9.docx]

References

1. Singh K, Fox JR, Brown RJ. Health anxiety and internet use: a thematic analysis. Cyberpsychology 2016;10(2):article 4. [doi: 10.5817/cp2016-2-4]
2. Akhtar M, Fatima T. Exploring cyberchondria and worry about health among individuals with no diagnosed medical condition. J Pak Med Assoc 2020;70(1):90-95. [doi: 10.5455/jpma.8682]
3. Fox S. Online health search 2006. Pew Research Center. 2006. URL: https://www.pewresearch.org/internet/2006/10/29/online-health-search-2006/ [accessed 2016-02-12]
4. Muse K, McManus F, Leung C, Meghreblian B, Williams JM. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the internet. J Anxiety Disord 2012 Jan;26:189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]
5. Berle D, Starcevic V, Khazaal Y, Viswasam K, Hede V, McMullan RD. Relationships between online health information seeking and psychopathology. Gen Hosp Psychiatry 2020;62:96-97. [doi: 10.1016/j.genhosppsych.2019.04.006] [Medline: 31029463]
6. White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. ACM Trans Inf Syst 2009;27(4):1-37. [doi: 10.1145/1629096.1629101]
7. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2016-02-12]
8. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008 Oct;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
9. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
10. Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J Oncol Pract 2016;8:737-744. [doi: 10.1200/jop.2015.010504]
11. Singh K, Brown RJ. Health-related internet habits and health anxiety in university students. Anxiety Stress Coping 2014;27(5):542-554. [doi: 10.1080/10615806.2014.888061] [Medline: 24467278]
12. Jungmann SM, Brand S, Kolb J, Witthöft M. Do Dr. Google and health apps have (Comparable) side effects? An experimental study. Clin Psychol Sci 2020;8(2):306-317. [doi: 10.1177/2167702619894904]
13. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
14. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related internet use. Expert Rev Neurother 2013;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
15. Bleichhardt G, Weck F. Kognitive Verhaltenstherapie bei Hypochondrie und Krankheitsangst. Heidelberg: Springer-Verlag; 2010.
16. Fergus TA. The Cyberchondria Severity Scale (CSS): an examination of structure and relations with health anxiety in a community sample. J Anxiety Disord 2014;28(6):504-510. [doi: 10.1016/j.janxdis.2014.05.006] [Medline: 24956357]
17. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord 2014;28:259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]
18. Norr AM, Oglesby ME, Raines AM, Macatee RJ, Allan NP, Schmidt NB. Relationships between cyberchondria and obsessive-compulsive symptom dimensions. Psychiatry Res 2015 Dec 15;230(2):441-446. [doi: 10.1016/j.psychres.2015.09.034] [Medline: 26429148]
19. Fergus TA, Russell LH. Does cyberchondria overlap with health anxiety and obsessive-compulsive symptoms? An examination of latent structure and scale interrelations. J Anxiety Disord 2016;38:88-94. [doi: 10.1016/j.janxdis.2016.01.009] [Medline: 26851749]
20. Fergus TA. Cyberchondria and intolerance of uncertainty: examining when individuals experience health anxiety in response to internet searches for medical information. Cyberpsychol Behav Soc Netw 2013;16(10):735-739. [doi: 10.1089/cyber.2012.0671] [Medline: 23992476]
21. Fergus TA. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria: a replication and extension examining dimensions of each construct. J Affect Disord 2015;184:305-309. [doi: 10.1016/j.jad.2015.06.017] [Medline: 26141807]
22. Fergus TA, Spada MM. Moving toward a metacognitive conceptualization of cyberchondria: examining the contribution of metacognitive beliefs, beliefs about rituals, and stop signals. J Anxiety Disord 2018;60:11-19. [doi: 10.1016/j.janxdis.2018.09.003] [Medline: 30317062]
23. Norr AM, Capron DW, Schmidt NB. Medical information seeking: impact on risk for anxiety psychopathology. J Behav Ther Exp Psychiatry 2014;45:402-407. [doi: 10.1016/j.jbtep.2014.04.003] [Medline: 24818986]

24. Fergus TA, Valentiner DP. Intolerance of uncertainty moderates the relationship between catastrophic health appraisals and health anxiety. Cogn Ther Res 2011;35:560-565. [doi: 10.1007/s10608-011-9392-9]
25. Norr AM, Albanese BJ, Oglesby ME, Allan NP, Schmidt NB. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. J Affect Disord 2015;174:64-69. [doi: 10.1016/j.jad.2014.11.023] [Medline: 25486275]
26. Brown RJ, Skelly N, Chew-Graham CA. Online health research and health anxiety: a systematic review and conceptual integration. Clin Psychol Sci Pract 2020;27(2):e12299. [doi: 10.1111/cpsp.12299]
27. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: systematic review and meta-analysis. J Affect Disord 2019;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
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;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
30. Fergus TA, Dolan SL. Problematic internet use and internet searches for medical information: the role of health anxiety. Cyberpsychol Behav Soc Netw 2014;17(12):761-765. [doi: 10.1089/cyber.2014.0169] [Medline: 25412398]
31. Fergus TA, Spada MM. Cyberchondria: examining relations with problematic internet use and metacognitive beliefs. Clin Psychol Psychother 2017;24(6):1322-1330. [doi: 10.1002/cpp.2102] [Medline: 28621035]
32. Starcevic V, Aboujaoude E. Cyberchondria, cyberbullying, cybersuicide, cybersex: "new" psychopathologies for the 21st century? World Psychiatry 2015;14(1):97-100 [FREE Full text] [doi: 10.1002/wps.20195] [Medline: 25655165]
33. Starcevic V. Cyberchondria: challenges of problematic online searches for health-related information. Psychother Psychosom 2017;86(3):129-133 [FREE Full text] [doi: 10.1159/000465525] [Medline: 28490037]
34. White RW, Horvitz E. Captions and biases in diagnostic search. ACM Trans Web 2013;7(4):1-28. [doi: 10.1145/2486040]
35. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2009;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
36. Eichenberg C, Schott M. Use of web-based health services in individuals with and without symptoms of hypochondria: survey study. J Med Internet Res 2019;21(6):e10980 [FREE Full text] [doi: 10.2196/10980] [Medline: 31199311]
37. Singh K, Brown RJ. From headache to tumour: an examination of health anxiety, health-related Internet use and 'query escalation'. J Health Psychol 2015;21(9):2008-2020. [doi: 10.1177/1359105315569620] [Medline: 25706333]
38. Norr AM, Allan NP, Boffa JW, Raines AM, Schmidt NB. Validation of the Cyberchondria Severity Scale (CSS): replication and extension with bifactor modeling. J Anxiety Disord 2015;31:58-64. [doi: 10.1016/j.janxdis.2015.02.001] [Medline: 25734759]
39. Barke A, Bleichhardt G, Rief W, Doering BK. The Cyberchondria Severity Scale (CSS): German validation and development of a short form. Int J Behav Med 2016;23(5):595-605. [doi: 10.1007/s12529-016-9549-8] [Medline: 26931780]
40. Bajcar B, Babiak J, Olchowska-Kotala A. Cyberchondria and its measurement. The Polish adaptation and psychometric properties of Cyberchondria Severity Scale CSS-PL. Psychiatr Pol 2019;53(1):49-60. [doi: 10.12740/pp/81799]
41. Gmelch M, Reinecker H. International scales for hypochondria. German-language adaptation of the Whiteley Index (WI) and the Illness Attitude Scales (IAS). Zeitschrift für Klinische Psychologie 2006;35(1):69-70 [FREE Full text]
42. Starcevic V, Baggio S, Berle D, Khazaal Y, Viswasam K. Cyberchondria and its relationships with related constructs: a network analysis. Psychiatr Q 2019;90(3):491-505. [doi: 10.1007/s11126-019-09640-5] [Medline: 31098922]
43. Henry J, Crawford J. 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;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
44. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychol Assess 2010 Mar;22:180-198. [doi: 10.1037/a0018260] [Medline: 20230164]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
46. Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
47. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Measur 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
48. Hautzinger M, Bailer K. Die Allgemeine Depressionsskala (ADS). Weinheim: Beltz Test Verlag; 1993.
49. Mewes R, Rief W, Brähler E, Martin A, Glaesmer H. Lower decision threshold for doctor visits as a predictor of health care use in somatoform disorders and in the general population. Gen Hosp Psychiatry 2008;30(4):349-355. [doi: 10.1016/j.genhosppsych.2008.04.007] [Medline: 18585539]

50. McElroy E, Kearney M, Touhey J, Evans J, Cooke Y, Shevlin M. The CSS-12: development and validation of a short-form version of the Cyberchondria Severity Scale. Cyberpsychol Behav Soc Netw 2019;22(5):330-335. [doi: 10.1089/cyber.2018.0624] [Medline: 31013440]
51. Jokić-Begić N, Mikac U, Čuržik D, Jokić C. The development and validation of the Short Cyberchondria Scale (SCS). J Psychopathol Behav Assess 2019;41(4):662-676. [doi: 10.1007/s10862-019-09744-z]
52. Shapira NA, Lessig MC, Goldsmith TD, Szabo ST, Lazoritz M, Gold MS, et al. Problematic internet use: proposed classification and diagnostic criteria. Depress Anxiety 2013;17(4):207-216. [doi: 10.1002/da.10094] [Medline: 12820176]
53. Spada MM. An overview of problematic internet use. Addict Behav 2014;39:3-6. [doi: 10.1016/j.addbeh.2013.09.007] [Medline: 24126206]
54. Meerkerk G, Van Den Eijnden RJ, Vermulst AA, Garretsen HF. The Compulsive Internet Use Scale (CIUS): some psychometric properties. Cyberpsychol Behav 2009;12(1):1-6. [doi: 10.1089/cpb.2008.0181] [Medline: 19072079]
55. Demetrovics Z, Szeredi B, Rózsa S. The three-factor model of Internet addiction: the development of the Problematic Internet Use Questionnaire. Behav Res Methods 2008;40(2):563-574. [doi: 10.3758/brm.40.2.563] [Medline: 18522068]
56. Wells A, Matthews G. Modelling cognition in emotional disorder: the S-REF model. Behav Res Ther 1996;34:881-888. [doi: 10.1016/s0005-7967(96)00050-2]
57. Bailey R, Wells A. Development and initial validation of a measure of metacognitive beliefs in health anxiety: the MCQ-HA. Psychiatry Res 2015;230:871-877. [doi: 10.1016/j.psychres.2015.11.035] [Medline: 26626951]
58. Warwick HM, Salkovskis PM. Hypochondriasis. Behav Res Ther 1990;28(2):105-117. [doi: 10.1016/0005-7967(90)90023-c]
59. Barsky AJ. Amplification, somatization, and the somatoform disorders. Psychosomatics 1992;33:28-34. [doi: 10.1016/s0033-3182(92)72018-0]
60. Newby JM, McElroy E. The impact of internet-delivered cognitive behavioural therapy for health anxiety on cyberchondria. J Anxiety Disord 2020;69:102150. [doi: 10.1016/j.janxdis.2019.102150] [Medline: 31739276]
61. Ferguson E. A taxometric analysis of health anxiety. Psychol Med 2008;39(2):277-285. [doi: 10.1017/s0033291708003322]
62. Longley SL, Broman-Fulks JJ, Calamari JE, Noyes R, Wade M, Orlando CM. A taxometric study of hypochondriasis symptoms. Behav Ther 2010;41(4):505-514. [doi: 10.1016/j.beth.2010.02.002] [Medline: 21035614]
63. Gibler RC, Mano KE, O'Bryan EM, Beadel JR, McLeish AC. The role of pain catastrophizing in cyberchondria among emerging adults. Psychol Health Med 2019;24(10):1267-1276. [doi: 10.1080/13548506.2019.1605087] [Medline: 30974950]
64. Mathes BM, Norr AM, Allan NP, Albanese BJ, Schmidt NB. Cyberchondria: overlap with health anxiety and unique relations with impairment, quality of life, and service utilization. Psychiat Res 2018;261:204-211. [doi: 10.1016/j.psychres.2018.01.002] [Medline: 29324396]
65. Brown RJ, Reuber M. Psychological and psychiatric aspects of psychogenic non-epileptic seizures (PNES): a systematic review. Clin Psychol Rev 2016;45:157-182. [doi: 10.1016/j.cpr.2016.01.003] [Medline: 27084446]
66. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
67. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174. [doi: 10.2307/2529310]
68. Salkovskis PM, Rimes KA, Warwick HM, Clark DM. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]
69. Longley SL, Watson D, Noyes R. Assessment of the hypochondriasis domain: the multidimensional inventory of hypochondriacal traits (MIHT). Psychol Assess 2005;17:3-14. [doi: 10.1037/1040-3590.17.1.3] [Medline: 15769224]
70. National Research Council. Combining Information: Statistical Issues and Opportunities for Research. Washington, D.C., United States: National Academy Press; 1992.
71. Schulze R. Current methods for meta-analysis. Zeitschrift für Psychologie / J Psychol 2007;215:90-103. [doi: 10.1027/0044-3409.215.2.90]
72. Hedges LV. Meta-analysis. J Edu Stat 1992;17(4):279-296. [doi: 10.3102/10769986017004279]
73. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychol Methods 1998;3:486-504. [doi: 10.1037/1082-989x.3.4.486]
74. Hunter JE, Schmidt FL. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (1st Ed). Thousand Oaks, California: SAGE Publications; 1990:1-592.
75. Hunter JE, Schmidt FL. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (2nd Ed). Thousand Oaks, California: SAGE Publications; 2004:1-616.
76. Viechtbauer W. metafor: meta-analysis package for R. R Package Version 1. URL: https://cran.r-project.org/web/packages/metafor/index.html [accessed 2017-07-01]
77. Viechtbauer W. Conducting meta-analyses in with the package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
78. Schmidt FL, Hunter JE. Measurement error in psychological research: lessons from 26 research scenarios. Psychol Methods 1996;1:199-223. [doi: 10.1037/1082-989x.1.2.199]
79. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions (Vol. 4). Hoboken, New Jersey, United States: John Wiley & Sons; 2011.

80. Beesdo-Baum K, Jenjahn E, Höfler M, Lueken U, Becker ES, Hoyer J. Avoidance, safety behavior, and reassurance seeking in generalized anxiety disorder. Depress Anxiety 2012;29(11):948-957. [doi: 10.1002/da.21955] [Medline: 22581482]
81. Te Poel F, Baumgartner S, Hartmann T, Tanis M. The curious case of cyberchondria: a longitudinal study on the reciprocal relationship between health anxiety and online health information seeking. J Anxiety Disord 2016;43:32-40. [doi: 10.1016/j.janxdis.2016.07.009] [Medline: 27497667]
82. Doherty-Torstrick ER, Walton KE, Fallon BA. Cyberchondria: parsing health anxiety from online behavior. Psychosomatics 2016;57(4):390-400 [FREE Full text] [doi: 10.1016/j.psym.2016.02.002] [Medline: 27044514]
83. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Personal Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
84. Carleton RN. The intolerance of uncertainty construct in the context of anxiety disorders: theoretical and practical perspectives. Expert Rev Neurother 2012;12:937-947. [doi: 10.1586/ern.12.82] [Medline: 23002938]
85. Carleton RN, Norton MP, Asmundson GJ. Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. J Anxiety Disord 2007;21(1):105-117. [doi: 10.1016/j.janxdis.2006.03.014] [Medline: 16647833]
86. Fergus TA, Bardeen JR. Anxiety sensitivity and intolerance of uncertainty: evidence of incremental specificity in relation to health anxiety. Personal Individ Diff 2013;55:640-644. [doi: 10.1016/j.paid.2013.05.016]
87. Abramowitz JS, Olatunji BO, Deacon BJ. Health anxiety, hypochondriasis, and the anxiety disorders. Behav Ther 2007;38:86-94. [doi: 10.1016/j.beth.2006.05.001] [Medline: 17292697]
88. Taylor S, Zvolensky MJ, Cox BJ, Deacon B, Heimberg RG, Ledley DR, et al. Robust dimensions of anxiety sensitivity: development and initial validation of the Anxiety Sensitivity Index-3. Psychol Assess 2007;19:176-188. [doi: 10.1037/1040-3590.19.2.176] [Medline: 17563199]
89. Sullivan MJ, 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]
90. McManus F, Leung C, Muse K, Williams JM. Understanding 'cyberchondria': an interpretive phenomenological analysis of the purpose, methods and impact of seeking health information online for those with health anxiety. Cogn Behav Ther 2015;7:e21. [doi: 10.1017/s1754470x14000270]
91. White RW, Horvitz E. Experiences with web search on medical concerns and self diagnosis. AMIA Annu Symp Proc 2009;2009:696-700 [FREE Full text] [Medline: 20351943]
92. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
93. Lucock M, Morley S. The health anxiety questionnaire. Br J Health Psychol 1996;1(2):137-150. [doi: 10.1111/j.2044-8287.1996.tb00498.x]
94. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996;11(3):89-96. [doi: 10.1097/00004850-199606000-00015]
95. The Whoqol Group. The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998;46(12):1569-1585. [doi: 10.1016/s0277-9536(98)00009-4]
96. Barsky AJ, Wyshak G, Latham KS, Klerman GL. Hypochondriacal patients, their physicians, and their medical care. J Gen Intern Med 1991;6(5):413-419. [doi: 10.1007/bf02598162]
97. Owens KM, Asmundson GJ, Hadjistavropoulos T, Owens TJ. Attentional bias toward illness threat in individuals with elevated health anxiety. Cognit Ther Res 2004;28(1):57-66. [doi: 10.1023/b:cotr.0000016930.85884.29]
98. Jasper F, Witthöft M. Health anxiety and attentional bias: the time course of vigilance and avoidance in light of pictorial illness information. J Anxiety Disord 2011 Dec;25(8):1131-1138. [doi: 10.1016/j.janxdis.2011.08.004] [Medline: 21890316]
99. Marcus DK, Gurley JR, Marchi MM, Bauer C. Cognitive and perceptual variables in hypochondriasis and health anxiety: a systematic review. Clin Psychol Rev 2007;27(2):127-139. [doi: 10.1016/j.cpr.2006.09.003] [Medline: 17084495]
100. Gropalis M, Bleichhardt G, Hiller W, Witthöft M. Specificity and modifiability of cognitive biases in hypochondriasis. J Consult Clin Psychol 2013;81:558-565. [doi: 10.1037/a0028493] [Medline: 22563641]
101. Witthöft M, Kerstner T, Ofer J, Mier D, Rist F, Diener C, et al. Cognitive biases in pathological health anxiety. Clin Psychol Sci 2015;4:464-479. [doi: 10.1177/2167702615593474]

Abbreviations

CBT: cognitive-behavioral therapy
CSS: Cyberchondria Severity Scale
DOCS: Dimensional Obsessive-Compulsive Scale
HA: health anxiety
MCQ-HA: Metacognitions Questionnaire–Health Anxiety
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SHAI: Short Health Anxiety Inventory

Edited by R Kukafka; submitted 09.02.21; peer-reviewed by P Copanitsanou, M Ak; accepted 21.07.21; published 18.11.21.

Please cite as:
Schenkel SK, Jungmann SM, Gropalis M, Witthöft M
Conceptualizations of Cyberchondria and Relations to the Anxiety Spectrum: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(11):e27835
URL: https://www.jmir.org/2021/11/e27835
doi:10.2196/27835
PMID:34792473

Original Paper

Patients' and Health Care Workers' Perception of Migraine Images on the Internet: Cross-sectional Survey Study

Bianca Raffaelli[1], MD; Pia Kull[1]; Jasper Mecklenburg[1], MD; Lucas Hendrik Overeem[1], MSc; Elisabeth Storch[1]; Maria Terhart[1]; Lars Neeb[1], MD; Uwe Reuter[1,2], MD, MBA

[1]Department of Neurology, Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Universitätsmedizin Greifswald, Greifswald, Germany

Corresponding Author:
Bianca Raffaelli, MD
Department of Neurology
Charité - Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450660888
Email: bianca.raffaelli@charite.de

Abstract

Background: The representation of migraine in the media is stereotypical. Standard images of migraine attacks display stylish young women holding their head in a pain pose. This representation may contribute to the social stigmatization of patients with migraine.

Objective: We aimed to analyze how patients with migraine and health care workers perceive online images of migraine.

Methods: The study consisted of an anonymous web-based survey of patients with migraine at the Headache Center of Charité – Universitätsmedizin Berlin (migraine group) and employees and students at our university (health care group). A total of 10 frequently used Adobe Stock photos of migraine attacks were presented to the participants. Each photo was rated on a scale of 0% to 100% based on how closely it resembled a realistic migraine attack (realism score). Patients with migraine also indicated how much each photo corresponded to their own experience of migraine as a percentage (representation score). We calculated the mean realism and representation scores for all photos and conducted further analyses using the categories male or female models, younger or older models, and unilateral or bilateral pain pose.

Results: A total of 367 patients with migraine and 331 health care employees and students completed the survey. In both groups, the mean realism score was <50% (migraine group: 47.8%, SD 18.3%; health care group: 46.0%, SD 16.2%). Patients with migraine identified their own migraine experience in these photos to a lesser degree (mean representation score 44.4%, SD 19.8%; P<.001 when compared to the realism score). Patients and health care workers considered photos with male models to be more realistic than photos with females (P<.001) and photos with older models to be more realistic than those with younger people (P<.001). In the health care group only, a bilateral pain posture was deemed more realistic than a unilateral pose (P<.001).

Conclusions: Standard images of migraine attacks are considered only slightly or moderately realistic by patients and health care workers. Some characteristics perceived as more realistic such as male sex or older age are in contrast with migraine epidemiology. A more accurate representation of migraine in the media could help to raise awareness for migraine and reduce the associated stigma.

(J Med Internet Res 2021;23(11):e32707) doi:10.2196/32707

KEYWORDS

migraine; stigma; mass media; stock photos; advocacy; internet; perception; headache; pain; cross-sectional; survey; stereotype; media; awareness

Introduction

Migraine is one of the most common neurological diseases, with a prevalence of 15% in the general population and rising to over 25% in women of childbearing age [1-3]. Migraine causes significant limitations in quality of life and functioning [4]. It is the second most common cause of health impairment among nonfatal diseases worldwide, as shown by the years lived with disability measure [4]. Among individuals aged 15 and 49 years, migraine ranks first among the most disabling of diseases [5].

Despite the substantial impact on patients' lives, migraine burden is often underestimated [6]. Many patients feel that their symptoms are dismissed as insignificant [7,8]. Platitudes such as "everybody has headaches" or "it's just stress" are omnipresent in the lives of patients with migraine [7,8]. In a survey by Buse et al [9] , almost half of patients with chronic migraine had the impression that their partner did not believe in their disease. The invisibility of migraine can lead to frustration and stigmatization [7]. Out of fear of being doubted, some patients even hide their symptoms and do not seek treatment, which in turn can have a negative impact on the course of the disease [10,11].

The lack of acceptance of migraine as a real disease has historical roots. Until recently, patients with migraine were portrayed as frail women with weak nerves [12]. Although this representation originates from a cultural background different from the present times, these stereotypes continue to shape the common view of patients with this disease [12].

Currently, digital media, and especially the internet, have become an important source of information on health topics [13]. Portrayals of people with migraine in the media can provide an overview on how society currently sees these patients [14]. Most images resulting from the search term "migraine" show slim and stylish young women holding their temples with an expression of pain on their faces [14]. This trivializing and one-sided portrayal could contribute to the insufficient recognition of migraine-related burden and the growth of social stigma [14].

While this stereotypical representation has already raised concerns among experts [14], no study has assessed how the public perceives such images of migraine. In this study, we aimed to investigate the following questions: (1) do patients with migraine and nonaffected health care workers perceive such photos as realistic? and (2) can patients with migraine relate to these portrayals?

Methods

Study Design

This anonymous web-based survey was performed on the REDCap (Research Electronic Data Capture) platform. The link to participate in the survey was distributed among the following two groups:

1. The migraine group: patients at the Headache Center, Charité – Universitätsmedizin Berlin, with a diagnosis of migraine in 2020 per International Classification of Headache Disorders–3 (ICHD-3) criteria [15];
2. The health care group: employees and students at the medical school of Charité – Universitätsmedizin Berlin without migraine.

Patients with migraine received the link to participate via a letter in order to comply with data protection law, while the health care group was invited via email distribution lists and social media groups.

The survey structure is illustrated in Table 1. After the assessment of demographic, occupational, and migraine characteristics, 10 different photos of migraine attacks were presented to the participants on the screen. The participants were instructed to rate on a scale between 0% and 100% how much each picture corresponded to a realistic migraine attack. We defined this percentage value as the realism score. Patients with migraine then indicated how closely each image resembled their own migraine experience on the same 0%-to-100% scale. This score was named the representation score.

Table 1. Structure of the survey.

Section	Description
Study information	Written information about the study design and aim, as well as the data protection statement agreement. Subjects could download the study information to keep for their records.
Written consent form	In order to access the other questionnaires, participants must confirm that they are ≥18 years of age, that they are voluntarily participating in the survey, and that they agree to the publication of the study results in an anonymous form. Participants could download a consent form to keep for their records.
Demographic characteristics	Participants are asked about their gender, age, ethnic background, height, weight, and highest level of education.
Migraine information	Participants are asked whether they experienced migraine, if they have close family members or friends with migraine, and how they assess the impairment caused by migraine in the general population and in their own migraine experience (on a numerical analog scale from 0% to 100%).
Occupational characteristics	Participants are asked whether they work or study at the Charité – Universitätsmedizin Berlin, in which field, and if they have regular contact with patients with migraine at work.
Photos 1-10	Participants are asked to look at 10 photos of migraine attacks and rate how realistic each photo is. Patients with migraine are also asked how representative of their own migraine experience these photos are.

The photos with models were obtained from the stock photo website Adobe Stock (Adobe Inc) [16]. We purchased a commercial license for the use of the 10 photos in the survey. Image selection was based on the following criteria:

- Result of the search term "migraine";
- Sorting by the number of times the photos were downloaded;
- 7 females and 3 males (to match the epidemiological sex distribution of migraine);
- Only 1 person in the photo;
- No black-and-white images;
- No heavy editing or heavy filters and effects;
- Face is visible;
- Only 1 photo of each model;
- Person in the foreground (ie, the background does not take more than half of the image).

Outcomes and Objective

The primary outcomes of the study were the mean realism score for all photos in both groups and the mean representation score for all photos in patients with migraine.

The secondary outcomes were the mean realism and representation scores for the following categories of photos:

- Photos with female models (n=7);
- Photos with male models (n=3);
- Photos with a unilateral pain posture (ie, models holding one side of the head, n=6);
- Photos with a bilateral pain posture (ie, models holding both sides of the head, n=3);
- Photos with younger models (n=5);
- Photos with older models (n=4).

The allocation of the photos into each category was agreed upon unanimously by all authors of this paper. To differentiate between younger and older models, we focused on physical characteristics such as face wrinkles or hair color (ie, white or gray).

Statistical Analysis

The analysis comprises all participants who rated all 10 photos. Employees or students with self-reported migraine were excluded from the health care group.

Demographic and occupational characteristics, as well as realism and representation scores, were summarized with descriptive statistics (absolute frequencies and percentages for categorical variables and mean (SD) values for numerical variables).

We compared the primary and secondary outcome measures between the migraine and health care groups using independent *t* tests. The realism and representation scores were compared within the migraine group using paired-sample *t* tests.

We further assessed the correlation of the primary and secondary outcomes with sex, age, ethnicity, highest level of education, and occurrence of migraine among family and friends using Pearson correlation analyses. A 2-tailed *P* value ≤.05 was considered statistically significant. *P* values were corrected for multiple comparisons with the Bonferroni method.

Ethics Approval and Consent to Participate

The study was approved by the ethics committee of the Charité – Universitätsmedizin Berlin (EA1/213/20). Participants were required to provide electronic consent prior to completing the survey.

Results

Population

Between October 27, 2020, and January 15, 2021, 367 patients with migraine and 331 Charité employees and medical students completed the survey. The participant selection process is illustrated in Figure 1.

The majority of patients were female (n=318, 86.6%), with an average age of 45.3 (SD 12.7) years. Participants in the health care group were younger (mean 32.1, SD 11.1 years), but their gender and ethnic distribution was similar to that of the migraine group (Table 2).

Figure 1. Flowchart of participant selection in both groups.

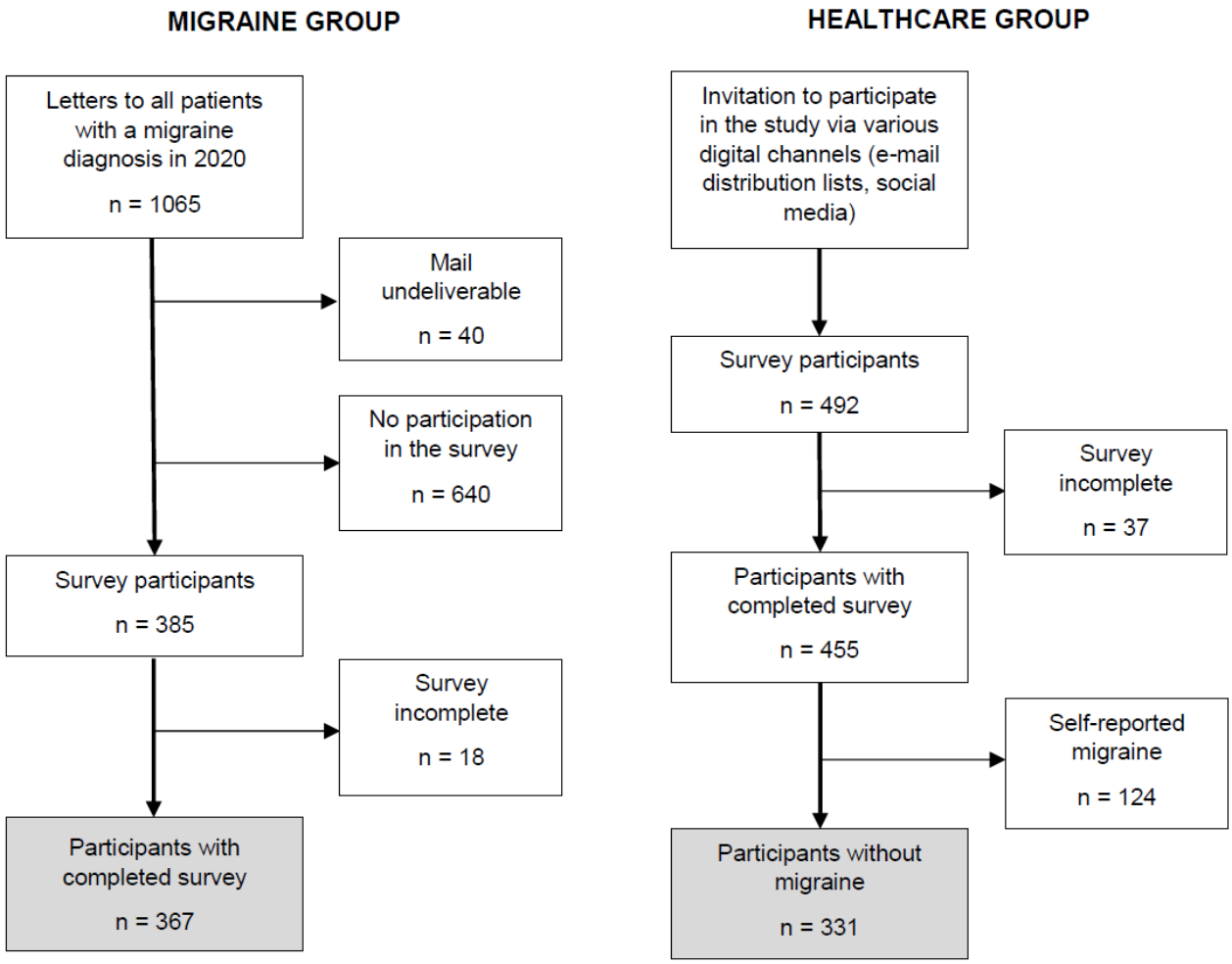


Table 2. Demographic and occupational characteristics of the survey participants.

Characteristic	Migraine group	Health care group
Age (years), mean (SD)	45.3 (12.7)	32.1 (11.1)
Female sex, n (%)	318 (86.6)	245 (73.9)
Northern or Central European descent, n (%)	314 (85.6)	283 (85.6)
Height (cm), mean (SD)	168.9 (8.1)	171.5 (9.7)
Weight (kg), mean (SD)	69.4 (14.6)	68.3 (13.3)
Highest level of education, n (%)		
University degree	154 (42.1)	118 (35.6)
High school diploma	63 (17.2)	155 (46.8)
Technical baccalaureate	29 (7.9)	8 (2.4)
Apprenticeship	60 (16.4)	20 (6.0)
Intermediate secondary school diploma (Realschulabschluss)	38 (10.4)	13 (3.9)
General secondary school diploma (Hauptschlussabschluss)	4 (1.1)	2 (0.6)
Other	19 (4.9)	15 (4.5)
Close friends or family members with migraine, n (%)	196 (53.4)	104 (31.4)
Health care workers' occupation, n (%)	—[a]	167 (50.5)
Physician	—	42 (25.1)
Nurse	—	33 (19.8)
Other medical professionals	—	27 (16.2)
Other nonmedical professionals	—	65 (38.9)
Health care students' study subject, n (%)	—	164 (49.5)
Human medicine	—	122 (74.4)
Dentistry	—	11 (6.7)
Other	—	31 (18.9)
Regular professional contact with patients with migraine, n (%)	—	68 (20.5)

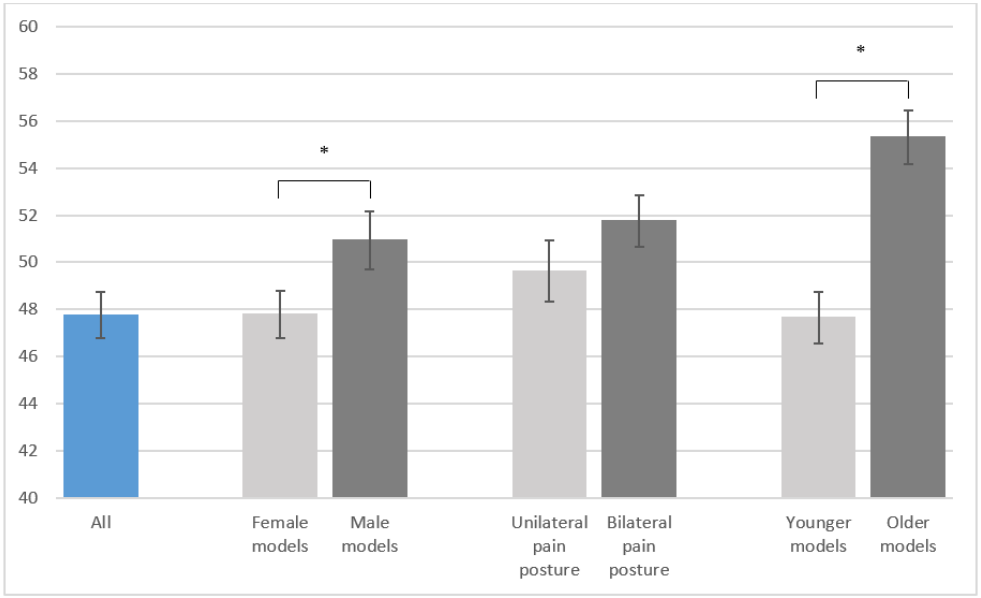
[a]Not applicable.

Patients With Migraine

Realism Scores

Among patients with migraine, the mean realism score for the 10 photos was 47.8% (SD 18.3%). Only 3 out of 10 photos had a mean realism score >50%. Photos with male models were considered more realistic than photos with female models (mean 51.0%, SD 22.4% vs mean 47.8%, SD 18.3%; *P*=.002). Patients rated images with older models as more realistic than those with younger models (mean 55.3%, SD 21.0% vs mean 47.7%, SD 20.1%; *P*<.001). Photos with unilateral and bilateral pain postures had similar realism scores (*P*>.99, Figure 2).

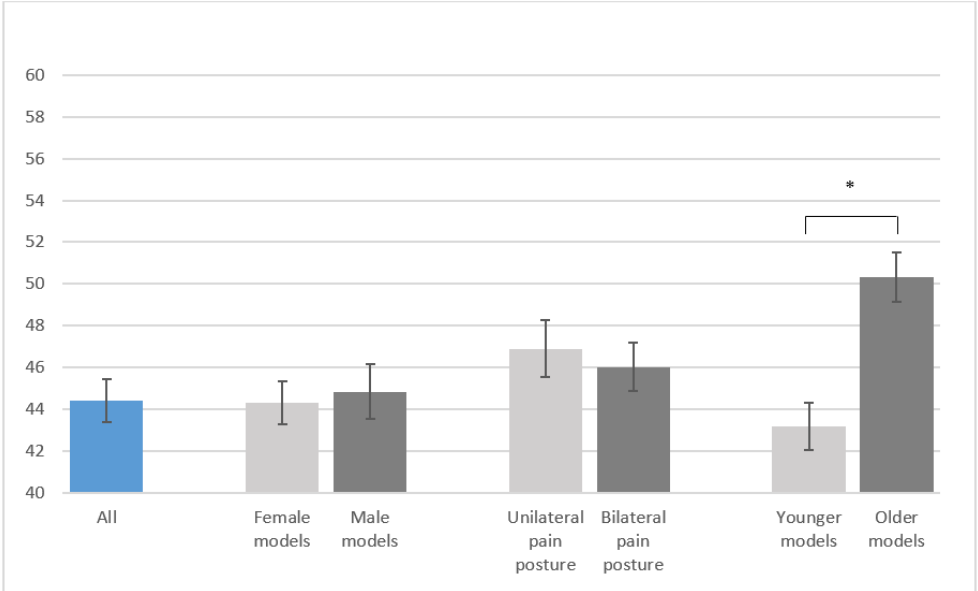
Figure 2. Realism scores of patients with migraine for all photos and different categories. Values are mean (SD), and the asterisk indicates *P*<.001.



Representation Scores

When asked how much the images corresponded to their own experience of migraine, the patients answered with a mean representation score of 44.4% (SD 19.8%). Photos with older models were considered more representative than photos with younger models (mean 50.3%, SD 22.7% vs mean 43.2%, SD 21.9%; *P*<.001). The gender of the models did not lead to significant rating differences in the representation score (*P*>.99). Photos with a unilateral pain posture had similar scores as did photos with a bilateral pain posture (*P*>.99, Figure 3).

Figure 3. Representation scores of patients with migraine for all photos and different categories. Values are mean (SD), and the asterisk indicates *P*<.001.



The mean representation score for all photos and in each category was significantly lower than the corresponding realism score (*P*<.001 for all categories).

There was a negative correlation between the highest level of education and both realism and representation scores: the higher the degree, the less realistic (*P*<.001, *r*=0.26) and representative (*P*<.001, *r*=0.29) the images were rated in all categories. Further analyses revealed a positive correlation between the patients' age and the realism (*P*=.047, *r*=0.11) and representation scores (*P*=.04, *r*=0.12) of images with older models. There was no correlation with the gender or ethnicity or with the occurrence of migraine among close friends or family members.

Health Care Workers and Students

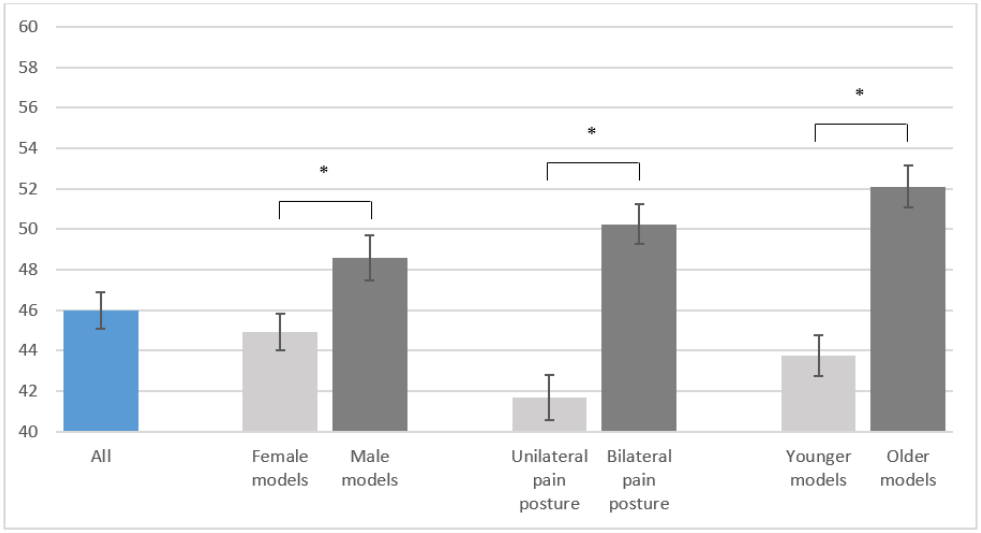
Realism Scores

In the health care group, the 10 photos had a mean realism score of 46.0% (SD 16.2%). Similar to the migraine group, only 3 photos were rated as >50% for realism. Photos with male models received higher scores than photos with female models (mean 48.6%, SD 20.2% vs mean 44.9%, SD 16.4%; *P*<.001). A bilateral pain posture was considered more realistic than a

unilateral pain posture (mean 50.2%, SD 18.1% vs mean 41.7%, SD 20.0%; *P*<.001). Photos with older models were rated higher than photos with younger models (mean 52.1%, SD 18.8% vs mean 43.8%, SD 18.3%; *P*<.001). Figure 4 shows the mean realism scores in the different categories.

There was no correlation to the gender, ethnicity, or age of the participant nor to regular contact with people with migraine (*P*>.99). Further, we detected a negative correlation with the highest level of education in this group (*P*=.004, *r*=0.16).

Figure 4. Realism scores of health care workers and students for all photos and different categories. Values are mean (SD), and the asterisk indicates *P*<.001.



Comparison Between Patients With Migraine and Health Care Workers and Students

Patients with migraine rated photos with a unilateral pain posture as significantly more realistic than the health care group (*P*<.001). Overall, the mean realism score for all photos did not differ between the two groups (*P*=.23). Both groups rated photos with male and older models as more realistic than photos with female and younger models. The mean realism scores for photos with females (*P*=.20), males (*P*>.99), younger models (*P*=.14), and older models (*P*=.53) were similar between groups.

Discussion

Patients with migraine and health care workers perceived commonly used stock photos of migraine attacks as slightly or moderately realistic. Patients identified their own migraine experience in these photos to an even lesser degree. Both groups rated photos with male and older models as more realistic than photos with female or younger ones. Among health care workers, a bilateral pain representation was considered more realistic than a unilateral pose.

The differences between media-based representations and the clinical reality have been described by Gvantseladze et al [14]. An analysis of the top 200 images under the search term “migraine” in 2 popular image-searching websites revealed that the majority of these images represented slim White females in a classic pain pose holding one or both temples [14]. The authors argued that this overrepresentation of ectomorph body types and stereotypical pain behaviors may contribute to the social stigmatization of patients with migraine [14]. Our results confirmed that not only experts but also patients doubt the realism of these images. In line with these findings, the Coalition for Headache and Migraine Patients (CHAMP) stated that media representation is often unrealistic and unlikely to display the severity of migraine [17]. The difference between the standard migraine representation and actual migraine behavior may result in the minimization of symptoms and misunderstanding of people living with this disease.

Such an example was illustrated by a social media trend from 2018, in which models and influencers published photos of themselves in the so-called “migraine pose,” touching one side of their face [18]. The fashion magazine *Elle USA* stated that this “flattering” pose “tightens the face, makes your cheekbones look more prominent, and lifts the brows” [18]. The use of the term migraine to name a glamorous pose indicates a lack of public acceptance for migraine as an extremely burdening condition. In line with the CHAMP Image Guide [17], our results support the need for a more accurate portrayal of migraine attacks. A better representation could include migraine symptoms other than headache, such as photophobia, nausea, or cognitive impairment [17]. A more diverse depiction could also help to move away from the classic temporal headache as the only accepted form of migraine pain. For example, a large proportion of migraine patients also have neck pain during attacks [19], which is almost never displayed as a feature of migraine [20].

Hospital employees and medical students shared the patients’ critical view of these images. This selected population was able to recognize that this stereotypical representation does not substantially correspond to reality. A more realistic representation of migraine attacks could also have a positive impact on patients’ treatment. The process toward effective migraine therapy is often lengthy and difficult [10]. The inaccurate, yet commonly accepted, representation of attacks could lead to a delay in the recognition and diagnosis of migraine, if patients experiencing an attack do not resemble

these common depictions [10]. Giving visibility to migraine in all its facets could therefore alleviate not only social stigmatization but also the therapeutic burden.

Migraine patients and the health care group rated the laterality of the headache pose differently. A unilateral representation was perceived as more realistic by the patients. Unilaterality is one of the key migraine characteristics according to the ICHD-3 but is not a mandatory criterion for migraine diagnosis [15]. Bilateral pain occurs frequently and, in older patients, it is even more common than strict unilateral attacks [21]. Patients with migraine at a tertiary headache center like ours might be better educated about the typical characteristics of migraine and therefore rate unilaterality as a more realistic migraine feature than the nonaffected group.

Migraine prevalence is 3 times higher in women than in men [3]. However, photos with male models were considered significantly more realistic than those with female models in this analysis, regardless of the rater's gender. This observation fits in well with the literature on gender bias, according to which pain disorders in women are taken less seriously than in men [22,23]. Pain expressions of females with chronic pain are underestimated compared to males, and women's pain is considered less severe [24]. In the health care system, women are less likely to receive pain medication than men [25]. On the contrary, psychosocial treatments are more often recommended to female patients experiencing pain [26]. Women with pain diseases are frequently met with skepticism and have to struggle to be believed, which might lead to shame and frustration [27]. If the woman is physically attractive, the credibility of her pain is even lower, which might be applicable to our photo models [28].

Similar considerations may apply to younger patients, especially if female. Young people with pain are often perceived as less ill, based on their healthy physical appearance [29]. This might explain why photos with older models were perceived as more realistic in our survey. This is in contrast with the epidemiology of migraine, which shows a prevalence peak during young adulthood [30].

Finally, our analysis showed that less educated people rated migraine stock images as more realistic than participants with a high level of education. People with a lower education are more likely to be influenced by the media [31,32]. Therefore, it is possible that they are accustomed to this type of migraine representation and do not question its realism. This high receptivity to the media could be useful for educational programs and campaigns to raise awareness for migraine and convey a more accurate and realistic representation.

This is the first study to analyze the perception of commonly used migraine images in a large cohort of patients with migraine and health care workers. Patients with migraine were selected directly from our Headache Center, which ensured a correct diagnosis. Due to data protection regulations, patients could not be contacted by email or telephone. Given that the only possible way to contact the patient was via mail, the response rate of over 30% is within the normal range of response.

A limitation of the study is that participants completed the survey anonymously online without supervision, which might have a negative impact on data reliability. Some biases may have affected our findings: people with a pre-existing awareness of this topic might have participated to a higher extent in the survey; this applies to younger people who are frequent users of the internet as well. We also divided the photos in six categories, but not all image characteristics were taken into account. For example, the outfit of the models, the surrounding environment, or the severity of the expression of pain were not considered and may represent confounding factors. In addition, only health care workers and students were enrolled in the comparison group, which might not be entirely representative of the general population. The extension of the survey to other members of the public might provide further insights on the perception of migraine.

To conclude, the media representation of migraine was considered at best moderately realistic in our large cohort of patients with migraine and health care workers. The rating of male and older models as more realistic contradicts migraine epidemiology. A more truthful representation of migraine is needed in order to raise awareness of the burden of this disease and to reduce migraine-related social stigma.

Conflicts of Interest

None declared.

References

1. GBD 2017 US Neurological Disorders Collaborators, Feigin VL, Vos T, Alahdab F, Amit AML, Bärnighausen TW, et al. Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study. JAMA Neurol 2021 Feb 01;78(2):165-176 [FREE Full text] [doi: 10.1001/jamaneurol.2020.4152] [Medline: 33136137]
2. Bigal ME, Liberman JN, Lipton RB. Age-dependent prevalence and clinical features of migraine. Neurology 2006 Jul 25;67(2):246-251. [doi: 10.1212/01.wnl.0000225186.76323.69] [Medline: 16864816]
3. Allais G, Chiarle G, Sinigaglia S, Airola G, Schiapparelli P, Benedetto C. Gender-related differences in migraine. Neurol Sci 2020 Dec 26;41(Suppl 2):429-436 [FREE Full text] [doi: 10.1007/s10072-020-04643-8] [Medline: 32845494]
4. Stovner LJ, Nichols E, Steiner TJ, Abd-Allah F, Abdelalim A, Al-Raddadi RM, et al. Global, regional, and national burden of migraine and tension-type headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2018 Nov;17(11):954-976. [doi: 10.1016/S1474-4422(18)30322-3]

5. Steiner TJ, Stovner LJ, Vos T, Jensen R, Katsarava Z. Migraine is first cause of disability in under 50s: will health politicians now take notice? J Headache Pain 2018 Feb 21;19(1):17 [FREE Full text] [doi: 10.1186/s10194-018-0846-2] [Medline: 29468450]
6. Katsarava Z, Mania M, Lampl C, Herberhold J, Steiner TJ. Poor medical care for people with migraine in Europe - evidence from the Eurolight study. J Headache Pain 2018 Feb 01;19(1):10 [FREE Full text] [doi: 10.1186/s10194-018-0839-1] [Medline: 29392600]
7. Cottrell C, Drew J, Waller S, Holroyd K, Brose J, O'Donnell F. Perceptions and needs of patients with migraine: a focus group study. Headache 2003 Apr;43(4):428-428. [doi: 10.1046/j.1526-4610.2003.03085_17.x]
8. Pearson C, Swindale R, Keighley P, McKinlay AR, Ridsdale L. Not Just a Headache: Qualitative Study About Web-Based Self-Presentation and Social Media Use by People With Migraine. J Med Internet Res 2019 Jun 19;21(6):e10479 [FREE Full text] [doi: 10.2196/10479] [Medline: 31219049]
9. Buse DC, Scher AI, Dodick DW, Reed ML, Fanning KM, Manack Adams A, et al. Impact of Migraine on the Family: Perspectives of People With Migraine and Their Spouse/Domestic Partner in the CaMEO Study. Mayo Clin Proc 2016 Apr 25;91(5):596-611 [FREE Full text] [doi: 10.1016/j.mayocp.2016.02.013] [Medline: 27132088]
10. Rutberg S, Öhrling K. Migraine--more than a headache: women's experiences of living with migraine. Disabil Rehabil 2012 Oct 10;34(4):329-336 [FREE Full text] [doi: 10.3109/09638288.2011.607211] [Medline: 21981545]
11. Lonardi C. The passing dilemma in socially invisible diseases: narratives on chronic headache. Soc Sci Med 2007 Oct;65(8):1619-1629. [doi: 10.1016/j.socscimed.2007.07.007] [Medline: 17716794]
12. Foxhall K. Wellcome Trust-Funded Monographs and Book Chapters. In: Migraine: A History. Baltimore, MD: Johns Hopkins University Press; 2019.
13. Sassenberg K. Digital media as laypeople's source of information about the environment and health. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2017 Jun;60(6):649-655. [doi: 10.1007/s00103-017-2549-2] [Medline: 28447133]
14. Gvantseladze K, Do TP, Hansen JM, Shapiro RE, Ashina M. The Stereotypical Image of a Person With Migraine According to Mass Media. Headache 2020 Jul 27;60(7):1465-1471. [doi: 10.1111/head.13846] [Medline: 32459017]
15. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. Cephalalgia 2018 Jan 25;38(1):1-211. [doi: 10.1177/0333102417738202] [Medline: 29368949]
16. Adobe Stock. 2021. URL: https://stock.adobe.com/de/ [accessed 2021-08-29]
17. Guides: Headache and Migraine Disease Language and Image Guide. Coalition for Headache and Migraine Patients (CHAMP). 2018. URL: https://headachemigraine.org/headache-and-migraine-disease-language-and-image-guide/ [accessed 2021-07-24]
18. Rodulfo K. All The Pretty People On Instagram Are Doing The Headache Pose. ELLE. 2018 Jun 26. URL: https://www.elle.com/beauty/a21947206/what-is-headache-pose-instagram/ [accessed 2021-07-24]
19. Ashina S, Bendtsen L, Lyngberg AC, Lipton RB, Hajiyeva N, Jensen R. Prevalence of neck pain in migraine and tension-type headache: a population study. Cephalalgia 2015 Mar 22;35(3):211-219. [doi: 10.1177/0333102414535110] [Medline: 24853166]
20. Rota E, Zucco R, Guerzoni S, Cainazzo MM, Pini LA, Catarci T, et al. Migraine Awareness in Italy and the Myth of "Cervical Arthrosis". Headache 2020 Jan 26;60(1):81-89. [doi: 10.1111/head.13679] [Medline: 31559636]
21. Martins KM, Bordini CA, Bigal ME, Speciali JG. Migraine in the elderly: a comparison with migraine in young adults. Headache 2006 Feb;46(2):312-316. [doi: 10.1111/j.1526-4610.2006.00343.x] [Medline: 16492241]
22. Samulowitz A, Gremyr I, Eriksson E, Hensing G. "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. Pain Res Manag 2018;2018:1-14 [FREE Full text] [doi: 10.1155/2018/6358624] [Medline: 29682130]
23. Hoffmann DE, Tarzian AJ. The girl who cried pain: a bias against women in the treatment of pain. J Law Med Ethics 2001 Jan 01;29(1):13-27. [doi: 10.1111/j.1748-720x.2001.tb00037.x] [Medline: 11521267]
24. Zhang L, Losin EAR, Ashar YK, Koban L, Wager TD. Gender Biases in Estimation of Others' Pain. J Pain 2021 Sep;22(9):1048-1059 [FREE Full text] [doi: 10.1016/j.jpain.2021.03.001] [Medline: 33684539]
25. Chen E, Shofer F, Dean A, Hollander J, Baxt W, Robey J, et al. Gender disparity in analgesic treatment of emergency department patients with acute abdominal pain. Acad Emerg Med 2008 May;15(5):414-418 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00100.x] [Medline: 18439195]
26. Hirsh AT, Hollingshead NA, Matthias MS, Bair MJ, Kroenke K. The influence of patient sex, provider sex, and sexist attitudes on pain treatment decisions. J Pain 2014 May;15(5):551-559. [doi: 10.1016/j.jpain.2014.02.003] [Medline: 24576430]
27. Werner A, Isaksen LW, Malterud K. 'I am not the kind of woman who complains of everything': illness stories on self and shame in women with chronic pain. Soc Sci Med 2004 Sep;59(5):1035-1045. [doi: 10.1016/j.socscimed.2003.12.001] [Medline: 15186903]
28. LaChapelle DL, Lavoie S, Higgins NC, Hadjistavropoulos T. Attractiveness, diagnostic ambiguity, and disability cues impact perceptions of women with pain. Rehabil Psychol 2014 May;59(2):162-170. [doi: 10.1037/a0035894] [Medline: 24611920]

29. Werner A, Malterud K. It is hard work behaving as a credible patient: encounters between women with chronic pain and their doctors. Social Science & Medicine 2003 Oct;57(8):1409-1419. [doi: 10.1016/s0277-9536(02)00520-8]
30. Burch RC, Buse DC, Lipton RB. Migraine: Epidemiology, Burden, and Comorbidity. Neurol Clin 2019 Nov;37(4):631-649. [doi: 10.1016/j.ncl.2019.06.001] [Medline: 31563224]
31. McKay DL, Houser RF, Blumberg JB, Goldberg JP. Nutrition information sources vary with education level in a population of older adults. J Am Diet Assoc 2006 Jul;106(7):1108-1111. [doi: 10.1016/j.jada.2006.04.021] [Medline: 16815128]
32. Jilani HS, Pohlabeln H, Buchecker K, Gwozdz W, De Henauw S, Eiben G, IDEFICS consortium. Association between parental consumer attitudes with their children's sensory taste preferences as well as their food choice. PLoS One 2018 Aug 1;13(8):e0200413 [FREE Full text] [doi: 10.1371/journal.pone.0200413] [Medline: 30067786]

Abbreviations

CHAMP: Coalition for Headache and Migraine Patients
ICHD-3: International Classification of Headache Disorders–3
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 06.08.21; peer-reviewed by P Gazerani; comments to author 27.08.21; revised version received 29.08.21; accepted 20.09.21; published 12.11.21.

Please cite as:
Raffaelli B, Kull P, Mecklenburg J, Overeem LH, Storch E, Terhart M, Neeb L, Reuter U
Patients' and Health Care Workers' Perception of Migraine Images on the Internet: Cross-sectional Survey Study
J Med Internet Res 2021;23(11):e32707
URL: https://www.jmir.org/2021/11/e32707
doi:10.2196/32707
PMID:34766918

Original Paper

Understanding the Relationship Between Official and Social Information About Infectious Disease: Experimental Analysis

Elias Assaf[1,2], PhD; Robert M Bond[3], PhD; Skyler J Cranmer[4], PhD; Eloise E Kaizar[2], PhD; Lauren Ratliff Santoro[5], PhD; Susumu Shikano[6], PhD; David J Sivakoff[2], PhD

[1]Pharmacy Analytics and Consulting, Research Consulting, Humana, Louisville, KY, United States

[2]The Ohio State University, Columbus, OH, United States

[3]Department of Communication, The Ohio State University, Columbus, OH, United States

[4]Department of Political Science, The Ohio State University, Columbus, OH, United States

[5]The University of Texas at Dallas, Richardson, TX, United States

[6]University of Konstanz, Konstanz, Germany

Corresponding Author:
Lauren Ratliff Santoro, PhD
The University of Texas at Dallas
800 W Campbell Rd.
Richardson, TX, 75080
United States
Phone: 1 9728836241
Email: laurenratliffsantoro@utdallas.edu

Abstract

Background: Communicating official public health information about infectious diseases is complicated by the fact that individuals receive much of their information from their social contacts, either via interpersonal interaction or social media, which can be prone to bias and misconception.

Objective: This study aims to evaluate the effect of public health campaigns and the effect of socially communicated health information on learning about diseases simultaneously. Although extant literature addresses the effect of one source of information (official or social) or the other, it has not addressed the simultaneous interaction of official information (OI) and social information (SI) in an experimental setting.

Methods: We used a series of experiments that exposed participants to both OI and structured SI about the symptoms and spread of hepatitis C over a series of 10 rounds of computer-based interactions. Participants were randomly assigned to receive a high, low, or control intensity of OI and to receive accurate or inaccurate SI about the disease.

Results: A total of 195 participants consented to participate in the study. Of these respondents, 186 had complete responses across all ten experimental rounds, which corresponds to a 4.6% (9/195) nonresponse rate. The OI high intensity treatment increases learning over the control condition for all symptom and contagion questions when individuals have lower levels of baseline knowledge (all P values ≤.04). The accurate SI condition increased learning across experimental rounds over the inaccurate condition (all P values ≤.01). We find limited evidence of an interaction between official and SI about infectious diseases.

Conclusions: This project demonstrates that exposure to official public health information increases individuals' knowledge of the spread and symptoms of a disease. Socially shared information also facilitates the learning of accurate and inaccurate information, though to a lesser extent than exposure to OI. Although the effect of OI persists, preliminary results suggest that it can be degraded by persistent contradictory SI over time.

(J Med Internet Res 2021;23(11):e25287) doi:10.2196/25287

KEYWORDS

disease; social information; official information; network experiments

Introduction

During a contagious disease outbreak, public health campaigns provide people with relevant information, including symptoms and methods of transmission. The public's understanding is critical for people to know which behaviors they should avoid and whether they should seek medical attention. To this end, public health campaigns are led by federal, state, local, and other organizations, which we refer to as *official* sources of information. However, many people may not directly encounter these campaigns and instead rely on information provided by social contacts, perhaps via social media. Rumors [1], lack of understanding [2], mis and disinformation campaigns [3], and motivated reasoning [4], among other factors, may inhibit understanding of the disease and contribute to a disconnect between accurate official information (OI) and inaccurate socially shared information. Furthermore, the spread of inaccurate information may have important downstream effects beyond the impact of the disease itself, such as on mental health [5].

Public health campaigns have been shown to be broadly effective. A recent examination of the COVID-19 pandemic in Italy found that most survey respondents knew of and believed in OI related to the disease [6]. This is particularly important, because in many ways socially shared misinformation about the disease was spreading more quickly than the disease itself [7]. In addition, experimental work demonstrates that individuals become more trusting of scientific experts and political actors with relevant expertise when encountering anxiety inducing external threats, such as H1N1 [8]. However, in addition to OI from the government and relevant policy actors, public health researchers have long recognized that people receive information from other sources, including media and friends, and that effective campaigns integrate these strategies [9]. A key challenge is understanding the impact that campaigns may have not only through direct exposure but also via social interactions and the sharing of disease-related information via social media.

People often rely on information from others to inform their own beliefs, attitudes, and behaviors related to health [10,11]. Studies have shown the importance of social networks in predicting health-related behaviors [12,13]. However, previous studies have not addressed the relative impact of socially transmitted information with respect to officially communicated public health information [see 14,15]. Specifically, van der Meer and Jin [14] studied the effectiveness of corrections to misinformation about infectious diseases coming from government health agencies, the news media, and social peers and found that government health agencies and the news media are more successful in combating misinformation about the disease than social peers. Similarly, Vraga and Bode [15] also looked at the effectiveness of corrections coming from the Centers for Disease Control and Prevention (CDC) or a social peer and found that the CDC is more effective in reducing misperceptions. Thus, engineers of public health interventions have little knowledge of how messages interact with socially spreading information. To address this, we used an experiment to causally test the hypothesis that health-related beliefs are socially influenced by the beliefs that others share with them in their social networks.

There is reason to believe that official, accurate health information is in competition with socially circulating inaccurate information. Rumors and misinformation are likely to spread across many domains [16]; research has shown this during recent infectious disease outbreaks, including HIV [17], H1N1 influenza [18], severe acute respiratory syndrome [19], the 2014-2015 Ebola outbreaks [20], the Zika epidemic [21], and the COVID-19 pandemic [7]. This information may be inaccurate and misinform the public or inhibit reliance on OI. Examining how these two information sources interact provides a deeper understanding of how the information environment leads to beliefs about an infectious disease.

This paper examines the interaction between official and social information (SI) about the disease. We use the term *social information* throughout, although we acknowledge that socially shared information differs from OI in many ways and that SI may be best understood as socially shared beliefs. To what extent does accurate, official public health information contribute to the learning of disease symptoms and transmission? To what extent does socially shared information magnify or inhibit learning? We used evidence from a randomized trial in a simulated network environment to shed light on these questions. In particular, we tested the hypotheses that participants who view relevant OI and those who view accurate SI will answer knowledge questions correctly more frequently than those with other informational experiences. We also examined whether these 2 effects reinforce each other. Finally, we tested the hypothesis that a positive teaching effect from a single viewing of OI can fade over time among those exposed to persistently inaccurate SI.

Theory and Expectations

Individuals' psychological dispositions interact with information from the external environment when making decisions [22]. However, individuals receive information from numerous sources. Individuals can receive information directly from the media, but most individuals are more likely to receive mediated information from opinion leaders in their social networks [23]. In today's digital age, this *social communication* frequently comes via the internet, including social networking sites. On the internet, owing to the ease of digital publishing, everyone has the ability to become an opinion leader [24]. Individuals are then tasked with determining which information is correct or incorrect and whether and how to update their beliefs based on the information they consume.

Public health information from official sources, where official sources include public health agencies such as the CDC and the World Health Organization (WHO), is likely to be accepted as factual by the public [15]. This is primarily because the public views these sources as credible, conveying both trustworthiness and expertise [14]. Specifically, work in public health demonstrates that government health agencies and news media organizations are viewed as more credible and are more successful than social peers in increasing perceptions of the severity of a public health crisis [14].

Although research has consistently found that individuals accept and view information from official sources as credible, the reality that individuals receive much of their information from social sources makes the overall information environment more challenging to navigate. Information from the social environment can contain correct, factual information about a disease, such as repetition of information shared by the CDC or WHO, or it can be misleading or even completely false. During the 2018 Ebola outbreak, for example, inaccurate socially supplied information about the disease was widespread, and individuals who believed in the misinformation were less likely to adopt preventative behaviors such as agreeing to be vaccinated and seeking formal care [25]. Still, SI can also be helpful in correcting misinformation, though to a lesser extent than information shared from official sources [14,15,26]. Given the complex nature of the information environments that people encounter, it is critical to understand how people reconcile official and SI, especially when it conflicts.

Several other features of the information environments that people encounter may affect how they process the messages they encounter. First, communication environments, particularly social media, are often dynamic in that there may be repeated interactions between users over time, and the use and effects of social media can be reciprocal [27]. In such situations, the effects of exposure to initial messages may degrade or be reinforced through repeated or subsequent interactions [28]. Therefore, understanding how conflicting social and OI signals affect people's attitudes over multiple interactions may provide a more nuanced understanding of how conflicting sources of information are processed.

In addition to being dynamic in time, communication on social media is often multidimensional even when it is limited to one overarching topic. For example, in the context of an infectious disease, two highly important dimensions of understanding the disease are the symptoms of the disease, so that a person may observe if they or others around them are likely to have contracted the disease, and how it spreads, so they know which actions are relatively risky or relatively safe in the context of the disease. Although it is important to know how people understand particular messages associated with either dimension in isolation, in many instances people process messages with multiple dimensions simultaneously, which may impact the effectiveness of a given message.

One of the primary features of social media is its ability to control self-presentation. Using the affordances of social media sites, users may selectively disclose information they wish their social contacts to know regarding their attitudes or beliefs [29] or may moderate their expressed attitudes because of their perception of the expectations of others [30]. In many instances, these processes are thought of in a research context as examples of social desirability bias, which are limitations of research designs or which research designs attempt to minimize to enable the observation of true attitudes or beliefs. Like many studies, our study cannot distinguish between sincerely and insincerely expressed beliefs. However, it is important to remember that even insincerely expressed attitudes or beliefs that are shared on social media platforms may still be taken at face value by others. Therefore, it is important to understand how attitudes and beliefs are expressed to others when people have the expectation that others will view those attitudes and beliefs. Because of the design of our study, we were able to examine the potential effects of the expression of beliefs, whether they were sincere or not.

Taken together, we expect that OI about a disease, that is, information received directly from official sources, will lead individuals to hold more factual beliefs about the disease. At the same time, socially shared information should lead to learning about disease, though to a lesser extent than OI. Specifically, we expect OI to be more effective in increasing learning about a disease when individuals are also exposed to accurate SI. On the other hand, we expect the effect of OI about a disease to persist but degrade as individuals are exposed to inaccurate SI over time.

Methods

Overview

To examine how official and SI interact in a dynamic information environment, we designed an experiment in which participants were exposed to both OI and structured SI over a series of ten rounds of computer-based interactions. The social connections between people are simulated with *bots* that are programmed to agree or disagree with key pieces of information in each round. These simulated alters allow us to control the social messaging received by each participant.

The research design for this study was reviewed and approved by the institutional review board of The Ohio State University (protocol #2014B0543).

Participants

Participants fluent in German were recruited from a participant research pool of a European university in 2016 and were compensated with a modest monetary incentive. Students signed up to participate in short sessions scheduled during the course of a single week. All students participating in the same session were randomly assigned to the same treatment combination, as described below.

The experiment was conducted using the oTree [31] software. After providing consent, the participants began by answering demographic and opinion questions. Participants were then asked to read a description of the study, which explained that they would be asked questions and claimed that their answers would be shared with other participants (as described below, actual participant responses were not shared with other participants). Specifically, participants were told that they were embedded in a social network with other student participants contemporaneously completing the experiment, that their responses would be shared with other participants, and that they would view the responses of 3 other participants. However, no network was actually created, and participants viewed responses preprogrammed to mimic those observed in a previous set of experiments that involved actual networked interactions.

Outcome Measures

Participants were then asked a battery of 14 true or false knowledge questions about hepatitis C, including six questions

about the modes of transmission and eight about the symptoms. The topic of each question is listed in Table 1, and Multimedia Appendix 1 [32,33] contains the complete knowledge question text.

Our novel 14 question instrument gauges participants' professed knowledge about hepatitis C. We reduce the impact of guessing by providing an uncertain answer, as well as true and false. The correct answers to all of the questions are presented in Table 1. Because participants learned about different topics related to hepatitis C at different times during the experiment (as described below), we did not combine participant responses across questions or across time points. Thus, we considered 14 separate outcome measures of professed knowledge, which correspond to indicators of a correct answer to each knowledge question. We did not distinguish between incorrect and uncertain responses in the analysis. Participant answers to the knowledge questions before exposure to treatments (as described in the *Design and Treatments* section below) provide a baseline to which we compare their future responses to the instrument as solicited throughout and at the end of the experiment.

Table 1. Summary of knowledge questions included in the experimental instrument, including the raw percent of participants who answered each question correctly at baseline, whether the question was subject to different social information, and the experimental round (if any) during which relevant official information was presented to participants in each intensity condition (high, low, and control). Note that infographics viewed by the low official information group were also viewed in the same round by participants in the high official information group.

Question topic	Correct answer	Correct at baseline (n=186), n (%)	Social manipulation	Round presented		
				High	Low	Control
Kissing	False	96 (51.6)	Different across conditions	1[a]	—[b]	—
Loss of appetite	True	81 (43.5)	Different across conditions	3[a]	3[c]	—
Headache	False	44 (23.7)	Different across conditions	—	—	—
Vomiting	True	78 (41.9)	Identical across conditions	3[a]	3[c]	—
Unprotected sex	True	154 (82.8)	Identical across conditions	5[a]	5[c]	—
Fever	True	127 (68.3)	Identical across conditions	7[a]	—	—
Fatigue	True	140 (75.3)	Identical across conditions	7[a]	—	—
Needle sharing	True	179 (96.2)	Identical across conditions	9[a]	9[c]	—
Breastfeeding	False	32 (17.2)	Identical across conditions	—	—	—
Diarrhea	False	35 (18.8)	Identical across conditions	—	—	—
Skin rash	False	43 (23.1)	Identical across conditions	—	—	—
Hair loss	False	83 (44.6)	Identical across conditions	—	—	—
Gym equipment	False	156 (83.9)	Identical across conditions	—	—	—
Tattoo equipment	True	165 (88.7)	Identical across conditions	—	—	—

[a]Official information treatment group.

[b]Not available.

[c]Social information treatment group.

We chose to measure knowledge about hepatitis C because it is a common disease with low public salience. Hepatitis C affects more than 3 million people in the United States (Department of Health and Human Services [34]), 14 million in Europe (WHO [35]), and approximately 71 million people worldwide (WHO [35]). We expected participants to have few entrenched beliefs and to find information about the disease novel. Summaries of the instrument responses reported in Table 1 show a range of baseline knowledge about hepatitis C topics, spanning from only 17.2% (32/186) reporting that hepatitis C is not spread via breastfeeding though an unsurprising 96.2% (179/186) reporting that hepatitis C is spread through sharing needles. The primary outcome measures capture changes in the proportion of correct answers within treatment groups, and the construction of these change score statistics is described in the *Statistical Analysis* section below.

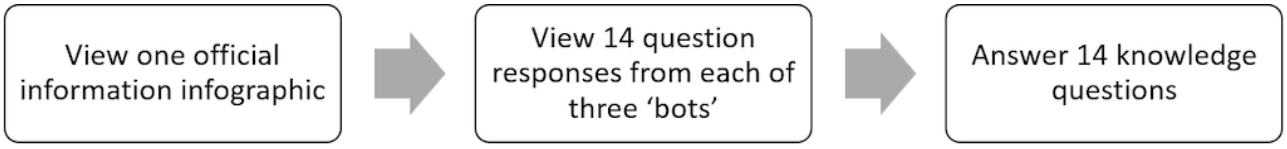
Design and Treatments

Ten experimental rounds were conducted at this point. We outline the overall structure of the rounds before describing how the treatments determine the nature of various components. Each round consisted of 3 parts, as shown in Figure 1. First, participants were exposed to an infographic intended to mimic information from a public health authority (*official* information). Some of this OI was related to knowledge questions about the modes of transmission and the symptoms of hepatitis C, and some were not. The infographics that were viewed in each round were determined by the treatment conditions, as described below (Table 1 for the infographic viewing schedule and question relevance and Multimedia Appendix 1 for the infographics themselves). Second, participants were exposed to ostensible SI from 3 other (purported) participants in the form of the participants' most recent answers to the 14 knowledge questions. A depiction of this SI is provided in Multimedia Appendix 1.

Third, participants were asked to answer the 14 knowledge questions.

We took care to create an environment in which participants believed that the observed SI was actually a response from other participants. First, participants completed the experiment in the same room as many other participants (sometimes participating in a different experiment). Second, our preprogrammed patterns of socially shared information were inspired by a pilot test of the experiment, such that relatively easy questions had more frequent accurate preprogrammed responses. Finally, we included time delays intended to mimic other participants that progressed more slowly. Although this SI does not mirror information shared from *friends*, it does correspond to information shared from casual connections across a social media platform.

Figure 1. Each round of the experiment consists of three parts. First, participants are exposed to official information, then they are exposed to ostensibly social information, and finally they are asked to respond to 14 knowledge questions.

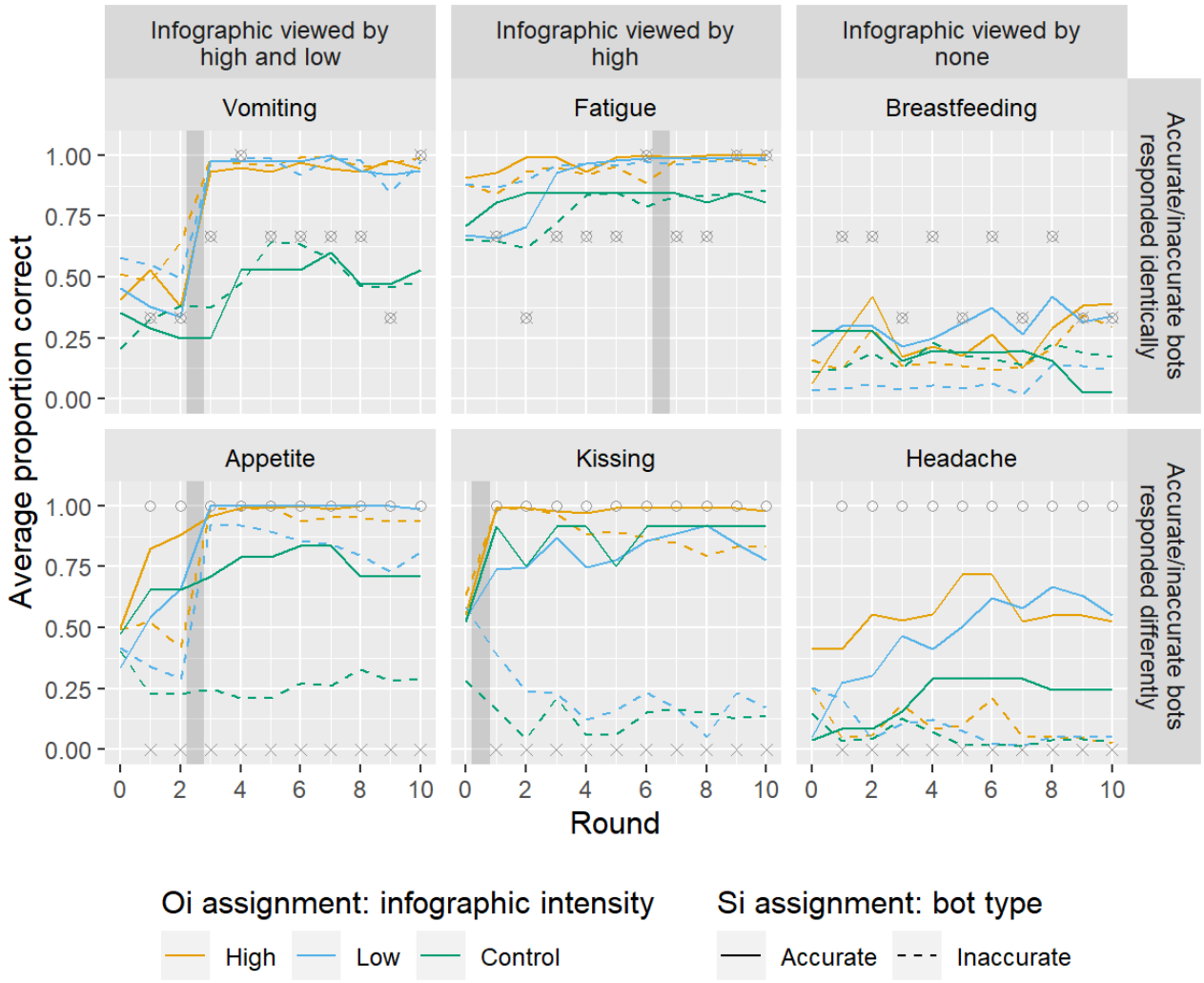


Participants were assigned by session via cluster randomization to one of three degrees (*high*, *low*, or *control*) of relevant OI and to either reinforcing (*accurate*) or contradicting (*inaccurate*) SI conditions. After combining extremely small sessions (Multimedia Appendix 1), the 22 resulting sessions were assigned to the cross-classified conditions. Four sessions were assigned to accurate SI and high degree of relevant OI (total n=38), 4 sessions were assigned to accurate SI and low degree of relevant OI (total n=26), and 3 sessions were assigned to accurate SI and control for OI (total n=17). Four sessions were assigned to inaccurate SI and high degree of relevant OI (total n=33), 4 sessions were assigned to inaccurate SI and low degree of relevant OI (total n=33), and 3 sessions were assigned to inaccurate SI and control for OI (total n=39).

Depending on their assigned OI condition, each participant could view 5, 3, or 0 OI infographics, each of which contained information directly relevant to 1 or 2 of the knowledge questions during the course of the experiment. Table 1 identifies the questions for which the participants were assigned to the *high*, *low*, or *control* categories. OI treatment conditions were presented directly relevant OI and the round during which this OI was presented. Infographics that conveyed information about hepatitis C unrelated to contagion or symptoms were presented in the remaining 5-10 rounds. The infographics are shown in Multimedia Appendix 1. All presented OI infographics contain accurate information.

The SI treatment conditions were intended to provide some participants with systematically accurate social influence and others with systematically inaccurate social influence, while maintaining the plausibility of real response sharing among participants. Thus, all participants viewed identical sequences of preprogrammed SI responses for 11 of the 14 knowledge questions. These preprogrammed responses were designed to mimic real participant responses to the same battery of questions deployed in a similar experiment. For the three remaining questions related to kissing, loss of appetite, and headache, participants assigned to the *accurate* SI condition were led to believe the three friends responded with false, true, and false (respectively, to the three questions) across all ten rounds, which are the correct responses. Those assigned to the *inaccurate* SI condition were led to believe the 3 friends responded with true, false, and true (respectively, to the 3 questions) across all ten rounds, which are the incorrect responses. Summaries of the displayed SI (bot) responses for 6 of the questions are depicted as circles and crosses in Figure 2. Multimedia Appendix 1 provides more information about the treatments and random assignment. At the end of each session, participants were provided with all correct factual information and were debriefed.

Figure 2. Proportion of correct responses for a representative subset of knowledge questions, averaged across sessions within the same cross-classified treatment assignments, as indicated by line color (official information [OI]) and type (social information [SI]). Direct interplay with treatment assignments differs across questions, as indicated by the columns (OI) and rows (SI) of panels. For OI, relevant infographics are viewed by the groups of participants indicated in the column heading in the round indicated by the vertical gray bar. For SI, the proportion of correct answers provided by bots are displayed as o and the proportion of incorrect answers provided by bots are displayed as x symbols. These symbols are always overlaid in the first row.



Statistical Analysis

For exploratory data analyses, we present question-specific longitudinal trends in the average proportion of correct responses across sessions stratified by treatment assignment. We colloquially refer to increases in the percentage of responses that are correct as learning, but recognize that these changes could also reflect changes in expressed attitudes rather than sincerely held beliefs. For formal statistical tests, we capture learning across rounds separately for each knowledge question via differences in average log odds ratios across sessions stratified by treatment assignment. This approach is similar to a random effects logistic regression where round is treated as a categorical factor but allows us to implement adjustments (Multimedia Appendix 1), which makes our calculations computationally feasible where classical random effects logistic regression is not.

Differences in learning over rounds and across treatment groups were summarized by subtracting the appropriate average log odds or log odds ratios. Most within-group summary statistics compared the log odds of a correct answer in a particular round to those at baseline, and some compared responses at round ten with those in the round where correct OI was viewed.

Finally, we compare learning across groups by taking the differences in within-group average log odds ratios across relevant treatment conditions. Large differences across treatment conditions imply that treatment causes differences in learning. We used permutation to approximate the exact reference distributions. To test the equality of learning across groups, we estimated 2-sided *P* values with the proportion of summary measures from 5000 permuted assignments that were further from zero than the summary measure based on the true assignment. Permutation-based tests rely on minimal assumptions but also typically have relatively low statistical power. Additional details regarding the statistical analyses are presented in Multimedia Appendix 1. A significance level of *P*=.05 was used for all tests. All analyses were conducted using R statistical software [32].

Results

Overview

A total of 195 participants across 23 sessions consented to participate; only data for the 186 who completed all ten rounds were included in the primary analyses and are described in Table 2. As noted in the table, 65.1% (121/186) of respondents were female, and their average age was 23 years. Overall, 56.5% (105/186) of respondents reported attending university for 4 or fewer semesters. Further demographic summaries as well as examination of missing data patterns and randomization balance are included in the Multimedia Appendix 1.

We tested four expectations linked to the relationship between accurate official public health information and accurate SI. These expectations imply that increases in correct responses across rounds (which we term *learning*) will differ across the 6 treatment groups in various ways for various questions.

Table 2. Baseline demographic and other characteristics for study completers (n=186).

Characteristic	Values
Age (years), mean (SD; range)	22.85 (2.44; 19 to 31)
Sex (female), n (%)	121 (65.1)
Parental socioeconomic status, self-assessed (n=180), n (%)	
Limited	35 (19.4)
Middle	95 (52.8)
Upper	50 (27.8)
Childhood environment, self-assessed (n=185), n (%)	
City	47 (25.4)
Small town	61 (32.9)
Suburbs	11 (5.9)
Country	66 (35.7)
Freshmen or sophomore class rank, n (%)	105 (56.5)
Political science major, n (%)	34 (18.3)
Political orientation score[a] (n=159), mean (SD; range)	−0.38 (0.75; −2 to 2)

[a]Political orientation was self-assessed on a 5-point scale from *very left* to *very right* and then recoded as integers from −2 to +2. Note that our European university recruitment pool provided little racial and ethnic diversity.

Learning From OI

We expect viewing OI to increase the learning of information about a disease, as previous literature demonstrates [6,8,15]. Those assigned to the high, low, and control intensity OI groups saw infographics relevant to seven, four, and zero questions. Thus, for 7 questions, we expect those who saw a relevant infographic (ie, those in higher intensity groups) to have greater learning for that question as compared with those who did not.

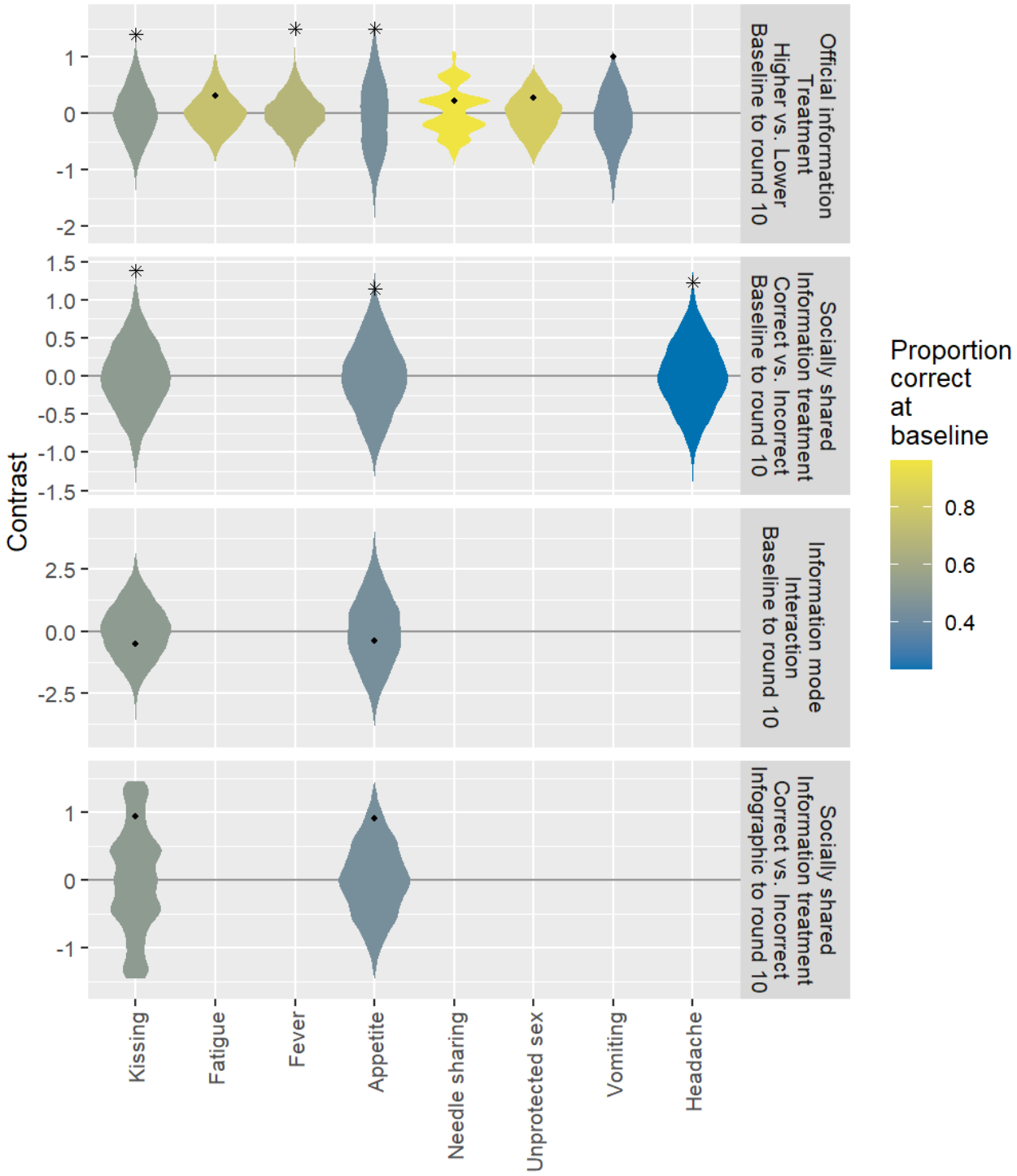
Visualization of data summaries highlights trends that may support our expectations. The lines in Figure 2 depict trends in the proportion of correct responses for each of the 6 groups (where color indicates OI treatment assignment and line type indicates SI treatment assignment) for 6 representative questions. Because all bots gave the same responses for the questions in the first row (as indicated by superimposed *x* and *o* symbols), these questions straightforwardly address the effect of OI. The leftmost plot in this row summarizes the responses to questions related to vomiting. The gray-shaded vertical bar indicates that a relevant infographic was shown to groups assigned to high-and low-intensity treatments before they answered the knowledge questions in round 3. For these 2 groups (yellow and blue lines), we see a dramatic increase in the proportion correct between rounds 2 and 3, suggesting that these groups learned from the relevant infographics. In contrast, we do not see similar learning among those assigned to the control intensity condition (green lines), suggesting that these groups did not learn from the irrelevant infographics they were shown.

For the question related to fatigue, we expected to see a similarly large learning in round 7 among the high intensity groups (and not among the low and control intensity groups who did not view a relevant infographic). However, differences in participant answers did not clearly follow this pattern, perhaps because of high baseline knowledge about this symptom or consistent longitudinal trends in bot responses. Further experiments are required to confirm this hypothesis. Finally, as expected, we see no substantial sustained learning about the potential contagion of breastfeeding in any group, as no group was shown a relevant infographic. Among the questions in the second row where the SI treatments differed, we also see substantial learning in rounds where relevant infographics are viewed, although these trends are complicated by differential socially shared information, as explained below.

Formal tests that compare groups assigned to view relevant infographics with those that did not are summarized for germane questions in the first row of Figure 3, where stars indicate statistically significant differences in the average log odds of a

correct question response across groups defined only by the OI treatment assignment. We found statistically significant differences in learning from baseline to round ten for all questions where the baseline knowledge was at most moderate (less than 75% correct), except for the question related to vomiting (P=.051). The results of pairwise comparisons across the assigned information treatment groups for all questions are included in Multimedia Appendix 1.

Figure 3. Observed (points) and permutation reference distributions (violins) for observed contrasts in average log odds of correct responses across treatment assignment groups (first two rows) and in difference-in-difference of average log odds (third row). Stars indicate permutation P≤.05.



Learning From Socially Shared Information

We expect (in) accurate socially shared information to affect learning of (in) correct information. For three of the questions (related to kissing, appetite, and headache), the (in) accurate bot responses were always (in) accurate; for the remaining 11 questions, all responses were identical across both types. Thus, for the 3 questions, we expect those assigned to the accurate bot treatment groups to learn more than those assigned to the inaccurate bot groups. The second row of Figure 2 shows the

response trends for these three questions, where the differential bot responses are indicated by 100% (3/3) and 0% (0/0) accurate responses for the accurate (*o*) and inaccurate (x) bots, respectively. Because no group was shown an infographic directly relevant to headache, this question (rightmost column) directly addresses the effect of SI. We see that across the ten rounds, those groups with whom accurate information is socially shared (solid lines) have positive learning, whereas those socially receiving inaccurate information (dashed lines) have flat or negative learning. In other questions, we see similar trends among groups that do not view relevant infographics (ie, the low and control intensity groups for the kissing question and the control group for the appetite question). The second row of Figure 3 shows that permutation tests confirm strong statistically

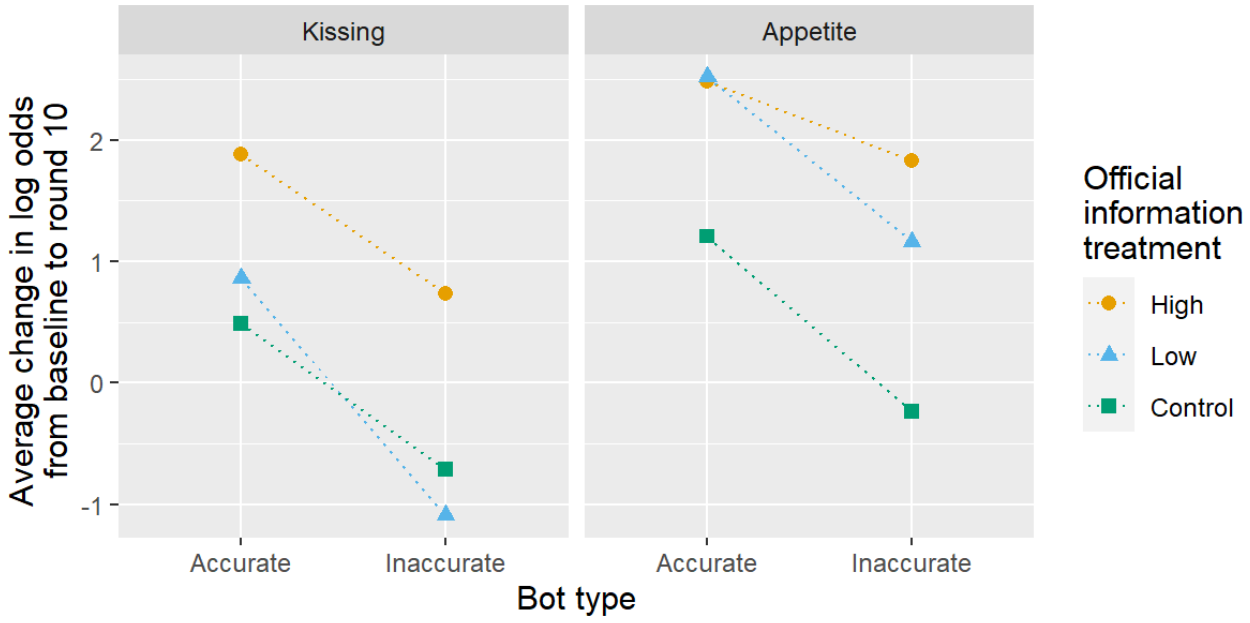
significant differences in learning from baseline to round 10 across the assigned bot types for all three questions, where the information shared by bots differed (all $P \leq .01$). Note that the participants had poor-to-moderate baseline knowledge for all three questions.

Interaction Between Information Sharing Modes

To the extent that official and SI interact, we expect OI to overwhelm SI. That is, we expect the effect of socially shared information to be smaller within groups who viewed germane infographics as compared with those who did not view germane infographics. Infographics relevant to appetite and kissing were shown to some groups, and the accurate and inaccurate bots differed in their responses to these questions. Thus, we focus on these two questions to explore information-mode interactions.

First, we consider the question related to appetite (lower left panel of Figure 2). Because all groups had similar knowledge at baseline, the effect of socially shared information can be approximated by the group differences in the average proportion correct at round 10. The effect of socially shared information is approximated by the difference between the corresponding solid and dashed lines. These differences are relatively small for the groups that viewed germane infographics (yellow and blue) and much larger for the groups that did not view germane infographics (green lines). This pattern is also present for the kissing question and seems to suggest that the effect of socially shared information is greater for groups that did not view germane infographics. However, this perception may be influenced by the boundary effects of the proportion scale. Figure 4 suggests that such differences disappeared after changing to the log odds scale. The log odds-based permutation tests (third row of Figure 3) confirm no strong evidence of an interaction between the modes of information sharing.

Figure 4. Treatment group-specific change in average log odds from baseline to round ten. Dotted lines connect pairs of groups with the same official information treatment group; parallel lines suggest no interaction between information modes.



Persistence of OI

Although we saw no global interactions among the effects of mode on learning from baseline to round 10, we explored a fourth, more targeted hypothesis that the effect of receiving official public health information persists but degrades in the face of contradictory SI. For the question related to appetite, we see that high- and low-information intensity groups achieved nearly unanimously correct responses after viewing the infographics in round 3 (all 4 yellow and blue lines in Figure 2). Of these, groups that are also assigned to socially receive accurate information (yellow and blue solid lines) retain a nearly unanimously correct response proportion at the end of the experiment. However, those groups with which inaccurate information is socially shared (yellow and blue dashed lines) tend to have small declines in the proportion of correct answers by round ten. We see similar trends for the high information intensity group (yellow lines) in answering the question related to kissing. The bottom row of Figure 3 confirms that these trends may be unusual relative to permuted treatment assignments, but the strength of evidence does not reach statistical significance (P=.29 and P=.09 for kissing and appetite, respectively).

It may also appear that there is more degradation in learning because of inaccurate SI for participants in the low OI group versus the high OI group (Figure 4, right panel, blue vs orange), and it is possible that participants receiving less OI overall might be more susceptible to social influence. However, we have only 2 groups with which to measure the variability of the persistence of learning because of OI; so, we did not attempt a formal comparison between these groups.

Discussion

Principal Findings

This work demonstrates that both official and SI influence people's reported understanding of infectious diseases. As expected from the previous literature, exposure to official public health information about hepatitis C increased the learning related to its spread and symptoms. Learning followed socially supplied information, whether accurate or inaccurate, though to a lesser extent. Trends in our data suggest that learning from OI is remarkably resilient even in the presence of persistent contradictory SI, although some modest degradation may be hidden by low statistical power. In an era in which official public health campaigns are frequently in competition with information shared on social media, this study provides some reason to be optimistic that public health campaigns may be able to overcome socially shared misperceptions.

Our work is consistent with previous work on misperceptions in the context of health, which finds that corrective information may inhibit false beliefs [14,15,36]. A web-based experiment investigating corrections to misinformation about a disease outbreak found that official sources of information were more effective at correcting misinformation than peers [14]. Our work examines the competing relationship between official and SI, rather than corrections to misinformation. However, we find that OI sources are much more effective at inducing changes in expressed beliefs, which is consistent with this previous work.

This study addresses the longitudinal effect of SI about diseases in tandem with official public health information in an experimental setting. Many studies of this type would include a one-shot intervention or follow up with participants after a short period. Here, we investigated how expressed beliefs update over multiple rounds and through multiple interactions with members of their experimentally constructed social networks. Because of this, we are able to provide a more nuanced understanding of competing messages in dynamic information environments.

It is important to note that our study used expressed beliefs about hepatitis C as the primary dependent variable. Of course, nearly any study of beliefs about disease relies on expressed beliefs in one way or another. However, because our study is premised on participants' understanding that they are viewing the responses of other participants who are completing the study contemporaneously, these expressed beliefs may be impacted by not only their own beliefs and any changes in them because of the experimental stimuli but also by social desirability bias that may impact their expressed beliefs [37]. Studies on the expression of belief in political falsehoods [38] have shown that the strategic expression of insincerely expressed beliefs is modest. However, the knowledge that you are sharing beliefs, whether sincerely held or not, means that these are beliefs that the participant expected others to see and experience. In this way, this work explores whether sincerely held beliefs are affected by official sources of information or not and whether the information environment becomes more clearly aligned with public health officials, reducing the degree of conflict between official and SI sources.

Although our work focuses on the context of a contagious disease, the results may contribute to our understanding of how social media may affect other aspects of health or may relate to other domains. Of course, OI may be in competition with (or be reinforced by) socially shared information across a number of domains, including other diseases, such as Zika [15], or aspects of health such as mental health [39]. Future work may wish to investigate whether and how socially circulating information affects beliefs in OI across other domains.

Strengths and Limitations

Our ability to control available information lends strong credibility to the causal interpretation of our results. Experiments on networks help bridge the gap between observational studies of people in their natural social environments and lab experiments that abstract away social influences.

Reliable extension of our conclusions to real-world situations relies on participants' interpretation of the information presented as legitimate. This hurdle may be easier to clear for OI than for SI. One major strength of our controlled experiment is that participants were embedded in an environment that promoted realism in a fictional social network.

This experiment has several important limitations, which we elaborate on here. First, the social networks we examined appeared to be among participants who were anonymous to one another within the setting of the experiment. Although this situation may be encountered on an anonymous message board or comment thread, it is possible that in real social networks, in which participants have strong social bonds and reputations, the effects of SI may be different. Future work may wish to investigate whether and how SI is transmitted in social networks among participants who are already connected to one another.

Second, our experiment limited SI responses to true, false, or unsure. In reality, people make intentional and nuanced attempts to convince others in the social sphere of their beliefs. It is difficult to envision a controlled experiment in which participants are allowed to communicate so freely and also maintains a degree of control that enables clear causal interpretation, but this should be addressed in future studies. We view this as a trade-off between tight control that enables us to make clear causal claims and less tight control that not only enables a greater variety of communication but also reduces the ability to interpret the social component in a clear causal fashion.

Third, the participants in our study were all similar in age. Previous work in a different context (politics) found that an official message with a reinforcing social component was more effective at changing behavior among older participants than

an official message without the social component [40]. It is possible that the limited interaction we find between social and OI owes to the age distribution of the participants of the study, the fact that the official and SI are contradictory, to the fact that our study examines health information, or to some other factor. Future work may examine contradicting official and SI in a more diverse sample to understand whether heterogeneous treatment effects exist.

Fourth, the study did not test for potential mediating and moderating variables that may explain how or to what degree official and SI affect beliefs. This limitation is related to, but distinct from, concerns about how anonymous networks may impact network effects. Mediating and moderating relationships may be a result of aspects of existing relationships, in which case anonymous studies would help to control for these effects but would also limit researchers' ability to examine them. However, it is also possible that mediating or moderating relationships are a result of aspects of relationships between people that develop rapidly, perhaps during the course of a few interactions. In this case, even in a short network experiment among anonymous individuals, mediating or moderating relationships may be examined. For example, the degree of trust in information sources may affect the degree to which participants respond to the beliefs shared by either official or SI sources. In this study, participants were informed that they would be connected to other anonymous participants; so, any interpersonal trust would have to be established during the course of the study. We believe it is unlikely that, in the context of this study, varying levels of trust would be created during the course of the study. However, in real-world situations in which people have ongoing relationships with those from whom they receive SI, trust and reputation are likely to impact how people process and respond to the messages shared by their social contacts. Future studies may wish to investigate trust, perhaps through experiments that enable researcher control over not only message content but also cues related to trust. We note that trustworthiness is one of multiple possible mediating or moderating variables that may enable a more nuanced understanding of the competing or reinforcing effects of social and OI.

Finally, the small number of experimental sessions limit the power of permutation-based statistical inference. The permutation-based approach used here makes very few assumptions and, therefore, enables us to make clear claims based on the statistical evidence that does not depend on common assumptions used in model-based analyses that may not hold for the data we collected. Future reanalyses based on longitudinal models may provide more nuanced, though more model-dependent results. In addition, reliance on college student volunteers and participant buy in to treatment legitimacy raises external validity concerns.

Nonetheless, our results provide some good news for sponsors of public health campaigns even in a time of overwhelming prevalence of SI with a wide range of accuracy. Consistent fact-based official messages can break through the SI noise.

Conclusions and Future Work

This experiment is one of multiple related projects examining social processes of information spread in networks and includes 2 additional experiments that address some of these limitations. First, human participants are randomly assigned to prespecified networks, which provides insight into how aspects of the network structure affect the diffusion of information about the disease. Second, participants share disease information with individuals in a real-world social network, providing a test of the external validity of our findings. Although this analysis focuses solely on the *bots* experiment, future work will incorporate the other 2 experiments. Although the studies of actual networks will enable the analysis of more complex network relationships, they will have a substantial trade-off in the degree of researcher control over what participants experience.

In the real world, our results present a puzzle for web-based health communities, particularly those that are moderated by professionals. In such communities, moderators may play a role similar to that of the official sources of information in our experiment [41,42]. In such communities, it may be possible for moderators to leverage network effects to broaden the impact of their official message. Web-based health communities often provide opportunities for users to connect in anonymous ways and must develop relationships with other users over time, similar to the design of our experiment. However, it is important to note that in such communities, professionals and moderators may have an outsized influence and the positive impact that they may have can lead to the propagation of misinformation just as it may lead to the spread of correct information [43,44]. Future work may wish to further investigate the interaction of official and SI or instances in which the line between them is not as clear, such as when a moderator or user of a web-based health community presents themselves as having expertise and other users must evaluate the user's credibility.

Our results underscore the importance of public health practitioners taking into account the effects of both OI sources, which may have a degree of control over, and SI sources, over which they may have limited control. We also show that OI is substantially more effective at promoting learning than socially supplied information, underscoring the importance of public health campaigns for inducing correct beliefs. Although we do not find evidence that official and SI interact, it is possible that the effects of such an interaction are small and nuanced, and we would nonetheless encourage practitioners to consider how OI may be transmitted through social ties.

Acknowledgments

This work was generously supported by the Alexander von Humboldt Fellowship for Experienced Researchers, the National Science Foundation (SES1357622, SES1461493, and SES1514750), and the National Institute of Drug Abuse at the National Institutes of Health (R34, DA043079). We thank the three anonymous reviewers for their insightful comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information concerning the design, analysis, results, examples of the treatments used, and primary dependent variable measurement.
[DOCX File , 3583 KB - jmir_v23i11e25287_app1.docx]

References

1. Friggeri A, Adamic L, Eckles D, Cheng J. Rumor cascades. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: International AAAI Conference on Web and Social Media; May 16, 2014; Ann Arbor, MI p. 101-110 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14559
2. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: An updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
3. Bode L, Vraga EK. See Something, say something: Correction of global health misinformation on social media. Health Commun 2018 Sep 16;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
4. Taber CS, Lodge M. Motivated skepticism in the evaluation of political beliefs. Am J Political Science 2006 Jul;50(3):755-769. [doi: 10.1111/j.1540-5907.2006.00214.x]
5. Chew AM, Ong R, Lei H, Rajendram M, Verma SK, Fung DS, et al. Digital health solutions for mental health disorders during COVID-19. Front Psychiatry 2020 Sep 9;11:582007 [FREE Full text] [doi: 10.3389/fpsyt.2020.582007] [Medline: 33033487]
6. Barari S, Caria S, Davola A, Falco P, Fetzer T, Fiorin S. Evaluating COVID19 Public Health Messaging in Italy: Self-reported compliance and growing mental health concerns. medRxiv 2020:1-19. [doi: 10.1101/2020.03.27.20042820]
7. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):1-2 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
8. Albertson B, Gadarian S. Anxious Politics: Democratic Citizenship in a Threatening World. New York, NY: Cambridge University Press; 2015.
9. Freimuth V, Linnan HW, Potter P. Communicating the threat of emerging infections to the public. Emerg Infect Dis 2000 Aug;6(4):337-347 [FREE Full text] [doi: 10.3201/eid0604.000403] [Medline: 10905966]
10. VALENTE TW, DAVIS RL. Accelerating the diffusion of innovations using opinion leaders. Ann Am Acad Polit Soc Sci 1999 Nov 01;566(1):55-67. [doi: 10.1177/000271629956600105]
11. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 02;334(6060):1269-1272. [doi: 10.1126/science.1207055] [Medline: 22144624]
12. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/nejmsa066082]
13. Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. Annu Rev Public Health 2007 Apr;28(1):69-93. [doi: 10.1146/annurev.publhealth.28.021406.144132] [Medline: 17222078]
14. van der Meer TG, Jin Y. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. Health Commun 2020 May 14;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
15. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Sci Commun 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]
16. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
17. Heller J. Rumors and Realities: Making sense of HIV/AIDS conspiracy narratives and contemporary legends. Am J Public Health 2015 Jan;105(1):43-50. [doi: 10.2105/ajph.2014.302284]
18. Shigemura J, Harada N, Tanichi M, Nagamine M, Shimizu K, Katsuda Y, et al. Rumor-related and exclusive behavior coverage in internet news reports following the 2009 H1N1 influenza outbreak in Japan. Disaster Med Public Health Prep 2015 Aug 17;9(4):459-463. [doi: 10.1017/dmp.2015.57] [Medline: 26186580]
19. Ding H. Rhetorics of alternative media in an emerging epidemic: SARS, censorship, and extra-institutional risk communication. Tech Commun Q 2009 Sep 17;18(4):327-350. [doi: 10.1080/10572250903149548]
20. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: A systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]
21. Carey JM, Chi V, Flynn DJ, Nyhan B, Zeitzoff T. The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in Brazil. Sci Adv 2020 Jan 29;6(5):eaaw7449 [FREE Full text] [doi: 10.1126/sciadv.aaw7449] [Medline: 32064329]

22. Boudon R, McPhee WN. Formal theories of mass behavior. Revue Française de Sociologie 1963 Oct;4(4):465. [doi: 10.2307/3319499]
23. Lazarsfeld P, Berelson F, Gaudet H. The Peoples Choice: How the Voter Makes Up His Mind in a Presidential Campaign. New York: Cambridge University Press; 1968.
24. Sundar S. The MAIN model: A heuristic approach to understanding technology effects on credibility. In: Metzger M, Flanagin A, editors. Digital Media, Youth, and Credibility. Cambridge, MA: MIT Press; 2008:73-100.
25. Vinck P, Pham P, Bindu K, Bedford J, Nilles E. Institutional trust and misinformation in the response to the 2018–19 Ebola outbreak in North Kivu, DR Congo: A population-based survey. Lancet Infect Dis 2019 May;19(5):529-536. [doi: 10.1016/S1473-3099(19)30063-5]
26. Vraga EK, Bode L. I do not believe you: How providing a source corrects health misperceptions across social media platforms. Inform Commun Soc 2017 Apr 19;21(10):1337-1353. [doi: 10.1080/1369118x.2017.1313883]
27. Wang Z, Tchernev JM, Solloway T. A dynamic longitudinal examination of social media use, needs, and gratifications among college students. Comput Hum Behav 2012 Sep;28(5):1829-1839. [doi: 10.1016/j.chb.2012.05.001]
28. Slater MD. Reinforcing Spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. Commun Theory 2007 Aug;17(3):281-303. [doi: 10.1111/j.1468-2885.2007.00296.x]
29. Bazarova NN, Choi YH. Self-Disclosure in Social Media: Extending the functional approach to disclosure motivations and characteristics on social network sites. J Commun 2014 Jun 19;64(4):635-657. [doi: 10.1111/jcom.12106]
30. Das S, Kramer A. Self-censorship on Facebook. In: Proceedings of the International AAAI Conference on Web and Social Media. 2021 Presented at: The International AAAI Conference on Web and Social Media; Jun 28, 2013; Cambridge, MA p. 120-127 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14412
31. Chen DL, Schonger M, Wickens C. oTree—An open-source platform for laboratory, online, and field experiments. J Behav Experiment Fin 2016 Mar;9:88-97. [doi: 10.1016/j.jbef.2015.12.001]
32. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL: https://www.r-project.org/ [accessed 2021-09-16]
33. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010 Aug 5;36(3) [FREE Full text]
34. Hepatitis C. United States Department of Health and Human Services. URL: https://www.hhs.gov/hepatitis/learn-about-viral-hepatitis/hepatitis-c-basics/index.html [accessed 2014-12-01]
35. Hepatitis. World Health Organization. URL: https://www.who.int/health-topics/hepatitis [accessed 2014-12-01]
36. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine 2015 Jan 09;33(3):459-464. [doi: 10.1016/j.vaccine.2014.11.017] [Medline: 25499651]
37. Krumpal I. Determinants of social desirability bias in sensitive surveys: A literature review. Qual Quant 2011 Nov 19;47(4):2025-2047. [doi: 10.1007/s11135-011-9640-9]
38. Berinsky AJ. Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. J Polit 2018 Jan;80(1):211-224. [doi: 10.1086/694258]
39. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: Systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
40. Bond RM, Settle JE, Fariss CJ, Jones JJ, Fowler JH. Social endorsement cues and political participation. Polit Commun 2016 Sep 28;34(2):261-281. [doi: 10.1080/10584609.2016.1226223]
41. 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]
42. Nath C, Huh J, Adupa AK, Jonnalagadda SR. Website sharing in online health communities: A descriptive analysis. J Med Internet Res 2016 Jan 13;18(1):e11 [FREE Full text] [doi: 10.2196/jmir.5237] [Medline: 26764193]
43. Ahmed W, Vidal-Alaball J, Downing J, Seguí F. COVID-19 and the 5G Conspiracy Theory: Social network analysis of twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
44. 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]

Abbreviations

CDC: Centers for Disease Control and Prevention
OI: official information
SI: social information
WHO: World Health Organization

Edited by R Kukafka; submitted 29.10.20; peer-reviewed by D Gunasekeran, B Narayan, S Tomczyk; comments to author 23.12.20; revised version received 01.03.21; accepted 12.08.21; published 23.11.21.

Please cite as:
Assaf E, Bond RM, Cranmer SJ, Kaizar EE, Ratliff Santoro L, Shikano S, Sivakoff DJ
Understanding the Relationship Between Official and Social Information About Infectious Disease: Experimental Analysis
J Med Internet Res 2021;23(11):e25287
URL: https://www.jmir.org/2021/11/e25287
doi:10.2196/25287
PMID:34817389

Original Paper

Mild Adverse Events of Sputnik V Vaccine in Russia: Social Media Content Analysis of Telegram via Deep Learning

Andrzej Jarynowski[1,2], DIS, MSc; Alexander Semenov[3,4], PhD; Mikołaj Kamiński, MD; Vitaly Belik[1], PhD

[1]System Modeling Group, Institute for Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Berlin, Germany

[2]Interdisciplinary Research Institute, Wrocław/Głogów, Poland

[3]Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, United States

[4]Center for Econometrics and Business Analytics, St. Petersburg State University, Saint Petersburg, Russian Federation

Corresponding Author:
Vitaly Belik, PhD
System Modeling Group
Institute for Veterinary Epidemiology and Biostatistics
Freie Universität Berlin
Königsweg 67
Berlin, 14163
Germany
Phone: 49 30 838 61129
Email: vitaly.belik@fu-berlin.de

Abstract

Background: There is a limited amount of data on the safety profile of the COVID-19 vector vaccine Gam-COVID-Vac (Sputnik V). Previous infodemiology studies showed that social media discourse could be analyzed to assess the most concerning adverse events (AE) caused by drugs.

Objective: We aimed to investigate mild AEs of Sputnik V based on a participatory trial conducted on Telegram in the Russian language. We compared AEs extracted from Telegram with other limited databases on Sputnik V and other COVID-19 vaccines. We explored symptom co-occurrence patterns and determined how counts of administered doses, age, gender, and sequence of shots could confound the reporting of AEs.

Methods: We collected a unique dataset consisting of 11,515 self-reported Sputnik V vaccine AEs posted on the Telegram group, and we utilized natural language processing methods to extract AEs. Specifically, we performed multilabel classifications using the deep neural language model Bidirectional Encoder Representations from Transformers (BERT) "DeepPavlov," which was pretrained on a Russian language corpus and applied to the Telegram messages. The resulting area under the curve score was 0.991. We chose symptom classes that represented the following AEs: fever, pain, chills, fatigue, nausea/vomiting, headache, insomnia, lymph node enlargement, erythema, pruritus, swelling, and diarrhea.

Results: Telegram users complained mostly about pain (5461/11,515, 47.43%), fever (5363/11,515, 46.57%), fatigue (3862/11,515, 33.54%), and headache (2855/11,515, 24.79%). Women reported more AEs than men (1.2-fold, $P<.001$). In addition, there were more AEs from the first dose than from the second dose (1.1-fold, $P<.001$), and the number of AEs decreased with age ($\beta=.05$ per year, $P<.001$). The results also showed that Sputnik V AEs were more similar to other vector vaccines (132 units) than with messenger RNA vaccines (241 units) according to the average Euclidean distance between the vectors of AE frequencies. Elderly Telegram users reported significantly more (5.6-fold on average) systemic AEs than their peers, according to the results of the phase 3 clinical trials published in *The Lancet*. However, the AEs reported in Telegram posts were consistent (Pearson correlation $r=0.94$, $P=.02$) with those reported in the Argentinian postmarketing AE registry.

Conclusions: After the Sputnik V vaccination, Russian Telegram users reported mostly pain, fever, and fatigue. The Sputnik V AE profile was comparable with other vector COVID-19 vaccines. Discussion on social media could provide meaningful information about the AE profile of novel vaccines.

(J Med Internet Res 2021;23(11):e30529) doi:10.2196/30529

KEYWORDS

adverse events; Sputnik V; Gam-COVID-Vac; social media; Telegram; COVID-19; Sars-CoV-2; deep learning; vaccine safety

Introduction

The current COVID-19 pandemic is one of the most critical global health problems. The main strategies for its mitigation involve both nonpharmaceutical interventions (eg, testing and contract tracing) and up-to-date anti-COVID-19 treatments. However, the most promising intervention has been vaccines that have effectively prevented severe COVID-19 outcomes. In addition to novel messenger RNA (mRNA) vaccines, vector vaccines have been developed. One of the first was Gam-COVID-Vac (Sputnik V), which is a viral, 2-dose, vector vaccine based on 2 human adenoviruses. Each dose contains a different vector: rAd26 and rAd5. This vaccine was developed by the Gamaleya Research Institute of Epidemiology and Microbiology. Sputnik V contains a gene that encodes SARS-CoV-2's spike (S) protein [1]. As of the time of this manuscript submission, 2 reports of clinical trials had been published. In the first study, phases 1/2 involved a total of 76 participant, who were included in the safety analysis [2]. The report on the phase 3 trial included detailed descriptions of serious and rare adverse events (AE) as well as mild AEs described in individuals [3] older than 60 years. The overall frequency of AEs was mentioned without complete characteristics of the safety profile, such as the co-occurrence of AEs. Mild AEs are common among all vaccines. Extensive fact sheets on AEs, as well as possible adverse reactions, were provided for vaccines trialed under the US Food and Drug Administration (FDA), UK Medicines and Healthcare products Regulatory Agency (MHRA), or EU European Medicines Agency (EMA), which was not the case with Sputnik V. As of April 17, 2021 (the end of the period for collecting data in our sample), 15,700,803 single doses of COVID-19 immunization had been administered in Russia [4]. The vast majority were of the Sputnik family (>95%), and the share of other vaccines was minimal (4.7% for EpiVacCorona and 0.1% for CoviVac) [5]. Moreover, the Russian Federation had signed contracts with dozens of countries to deliver 1.4 billion doses at less than €7 (US $8.13) per dose for international buyers [6]. Therefore, there is an emerging need to update the information on Sputnik V's safety profile using postmarketing surveillance. Because a registry of AEs after vaccination with Sputnik V is difficult to access, social media discourse may be an alternate source of information on AEs. The Sputnik vaccine gave rise to dubious situations in not only its safety profile but also other aspects [7].

An increasing number of studies has analyzed English-language social media in the context of vaccinations [8] or vaccine-prevented infectious disease [9]. However, only a few similar studies on Russian social media have been published [10,11]. Accounts of adverse reactions to drugs have been widely extracted from social media [12] in the context of mining consumer reviews on the internet [13]. To date, most of these studies processed data collected from Twitter [14-23]. Although social media platforms such as Twitter and Facebook are used in Russia, Telegram Messenger is ranked second in the Russian App Store, having 27 million active users in Russia [24]. Developed in Russia, this platform is much more popular than alternatives such as Twitter [25].

Most previous studies on social media vaccine discourse have focused on the personal beliefs of users. For example, Wang et al [26] developed a framework to detect vaccine AEs mentioned by Twitter users. However, to date, no study has analyzed social media discourse on nonsevere AEs in response to COVID-19 vaccines. In this study, we collected social media (a Telegram group in the Russian language) data to bridge the gap in information on the most prevalent AEs involving Sputnik V. We focused on the most common AEs and established which were the most prevalent, their co-occurrence, and their associations with users' characteristics [27]. Finally, we compared the AE profile of Sputnik V with those of other approved COVID-19 vaccines.

Methods

The dataset analyzed in our study was collected retrospectively from the Telegram group, "Sputnik_results" [28]. The data contained no personal information, and the analysis was performed according to the Terms of Service of the platform [29]. Our analysis was completely anonymous and performed in aggregated form. No possible harm to Telegram users was identified. Therefore, the study did not require ethical committee approval.

Data Description

Originally, Telegram aimed to provide secure communication (which is very important for post-Soviet societies [30]), but later, functionality was expanded; it added support for public channels, groups, video calls, and many other features [29]. Telegram groups may be public or private. If a group is public, it may be accessed via the Telegram search engine, and every user may read all its content. A main priority claimed by Telegram is security; users' data are not disclosed, and only the user's screen name and picture are shown to the public. The largest Telegram channels have millions of subscribers.

The description of the "Sputnik_results" [28] public group states that its main aim is to collect information on AEs regarding the Sputnik V vaccine. Telegram users may post a description of their symptoms. Moderators of the group oversee the messages and verify that they contain only descriptions of AEs; otherwise, the message is deleted. An example message is as follows: "М, 33 года. V1 24.01.21 через 12 часов темп 39, боль в руке (все плечо целиком, мышцы), заложенность носа, диарея. На следдень темп 38, боль в руке, заложенность носа. На третий день слабость, температура в норме" (translation: M, 33 years old. V1 24.01.21 after 12 hours, temp. 39, pain in the arm (the entire shoulder, muscles), nasal congestion, diarrhea. The next day, temp. 38, pain in the arm, nasal congestion. On the third day, weakness, temperature is normal).

In this study, we collected all messages from the "Sputnik_Results" group using Python Telegram Client telethon [31]. We saved only text messages that were posted in the group; users' personal details were not extracted. In total, we collected 18,833 messages. After filtering messages that contained only

pictures, 11,515 messages remained. The first message was sent on December 9, 2020, and the most recent message was sent on April 17, 2021. The dataset contained 25,660 unique lowercase words.

Adverse Event Classification

The gold standard used to identify adverse reactions is the MedDRA System Organ Class, which is applied in the European Union (EudraVigilance [32]), the United States (Vaccine Adverse Event Reporting System [VAERS] [33]), and the United Kingdom (MHRA Yellow Card scheme [34]). However, the system uses a specialized medical vocabulary. In our study, because users of social media communicated in colloquial language [12], we chose a simplified FDA classification system [35-37] that was subdivided into 2 groups: local reactions (ie, redness, swelling, and pain at the injection site) and systemic reactions (ie, fever, fatigue, headache, chills, nausea/vomiting, diarrhea, new or worsening muscle pain, and new or worsening joint pain). Moreover, muscle pain, joint pain, and pain at the injection site were categorized as a single class. However, we added the classes of pruritus, enlarged lymph nodes, and insomnia, which are common adverse reactions to anti-COVID-19 vaccines [38,39]. Insomnia was chosen due to its high frequency by simple keyword analysis on a sample of material from Telegram. The final list of 12 classes of symptoms of mild AEs, which were based on subjective experiences of a potential health issue, is provided in the Results section.

Labeling

We utilized the LabelStudio data labeling tool [40] to label the dataset. We randomly sampled 1000 messages in the dataset, which were labeled by 3 raters who were native Russian speakers. The raters labeled each occurrence of an AE in the messages, thus making the dataset suitable for named entity recognition tasks. Because of such labeling and the existence of different descriptions of the same AEs in multiple sentences, we augmented the dataset by splitting each message into sentences. The resulting dataset contained 4579 entities.

Model Architecture

Each message in our dataset could have included multiple AEs. We therefore adopted a multilabel text classification scheme. A formal definition of multilabel classification is as follows: Consider a dataset

where $x_i \in X$ is the i-th observed variable for the dataset of cardinality n, $y_i \in Y$ is the corresponding set of labels for the i-th element. Our goal was to learn a mapping $\hat{y}_j = f(x_j,\theta)$, where $\hat{y}_j$ is the set of predicted classes and θ is a vector of parameters. To find the vector of optimal parameters θ, we needed to minimize the loss function $L(y,\hat{y})$ between the actual and predicted classes. Multiple machine learning methods may be applied to support multilabel classification. In the case of artificial neural networks (ANNs), the activation function of the last layer of the ANN is set to be a sigmoid:

and binary cross-entropy loss is used. In this case, ANN will map the probability of each class to a value between 0 and 1, and each data item could be mapped to multiple classes.

Because of the recent success of ANNs, specifically transformers, in text analysis tasks, we adopted a deep Bidirectional Encoder Representations from Transformers (BERT) architecture to perform our multilabel classification task [41]. We utilized a pretrained BERT model for the Russian language DeepPavlov [42]. We tuned the last layer of the model, which consisted of 12 sigmoid neurons. As a baseline, we used a standard long short-term memory (LSTM) ANN, which consisted of embedding as the first layer and 1 LSTM layer (100 cells), dropout (P=.20), and a subsequent multilabel dense layer with sigmoid as the activation function.

Model Evaluation

We trained the BERT and LSTM models using a stratified k-fold validation scheme where k=5. Because the classes were imbalanced, we utilized an up-sampling strategy; that is, underrepresented classes were up-sampled in the training dataset. The testing set distribution was not modified. Table 1 displays the evaluation results. Precision and recall were calculated for both micro- and macro-averaged aggregations [43]. As shown in Table 1, precision and F1 scores were reported for thresholds equal to 0.5. We utilized a computer with a Tesla T4 GPU to train the models. Table 1 shows that BERT outperformed the LSTM model by a large margin. We therefore chose the BERT model and trained it on 95% of the data; in this case, it returned a micro-averaged accuracy of 0.94 and an area under the receiver operating characteristic (ROC) curve (AUC) score of 0.991.

Regarding gender, age, and dose number (if available), we used counts of corresponding abbreviations and regular expression matching because the administrators of the group had provided detailed instructions for the reporting of this information.

Table 1. Bidirectional Encoder Representations from Transformers (BERT) and long short-term memory (LSTM) model evaluation results.

Model	Micro-averaged aggregations			Macro-averaged aggregations	
	AUC[a], mean (SD)	Precision, mean (SD)	F1, mean (SD)	Precision, mean (SD)	F1, mean (SD)
LSTM	0.969 (0.002)	0.866 (0.024)	0.769 (0.033)	0.514 (0.048)	0.431 (0.042)
BERT	0.991 (0.002)	0.915 (0.016)	0.920 (0.002)	0.863 (0.025)	0.858 (0.006)

[a]AUC: area under the curve.

Analysis of AE

To evaluate the time relationship between the number of reports and vaccination volume, a univariate linear regression coefficient was calculated. Because the number of reports (P<.001) and vaccination volume (P<.001) failed to be normally distributed based on the Shapiro-Wilk test, a Spearman correlation was calculated. Because the number of AEs failed to be normally distributed based on the Shapiro-Wilk test (P<.001), the difference between the 2 groups was analyzed with a Mann-Whitney U test. To compare frequencies of AEs between 2 samples of AEs, a Fisher test was applied. To compare the frequencies of 2 vectors of AEs, the normality was checked with the Shapiro-Wilk test, and Pearson correlations could be calculated (P=.10 and P=.07, respectively, comparing Telegram with the Argentinian Registry; P=.13 and P=.34, respectively, comparing Telegram with the Moscow trial). Community detection was conducted to evaluate the internal structure (co-occurrence) of AEs in the network representation.

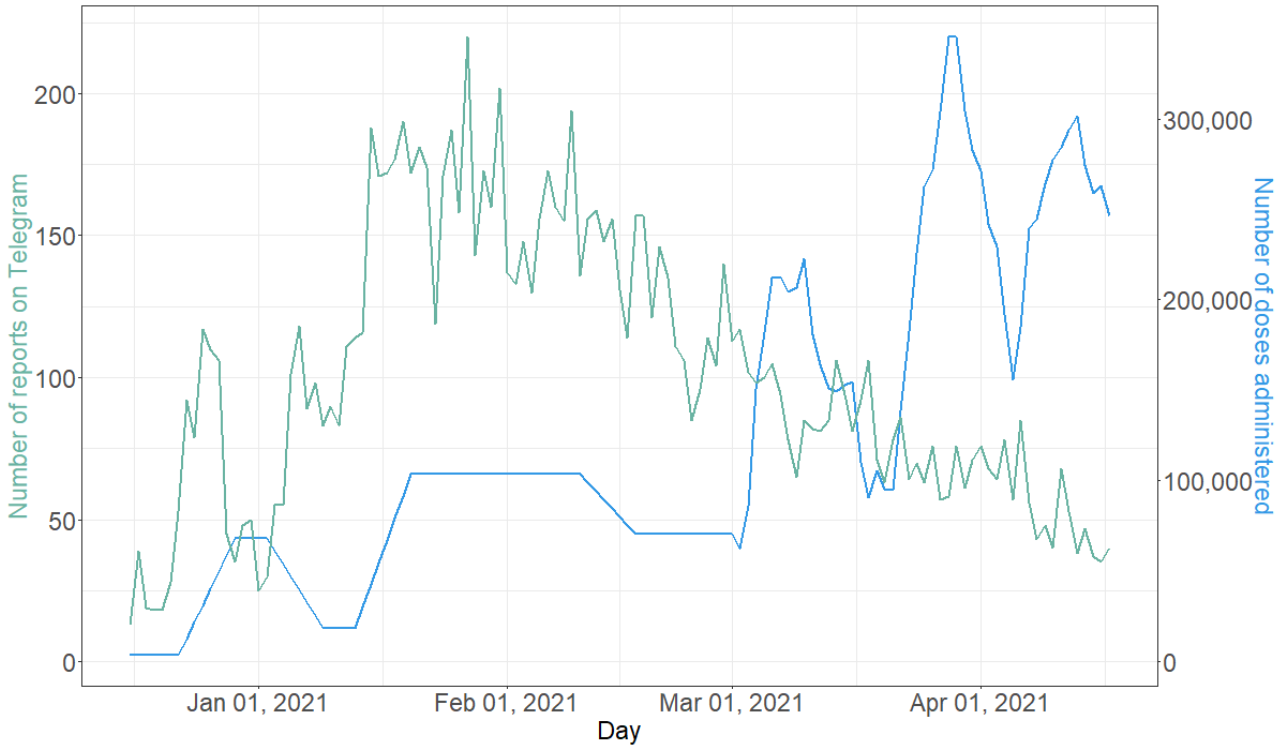
Results

Reactogenicity assessment based on opt-in civic surveillance was performed to obtain results of clinical importance (similar to endpoints in trials).

Temporal Dynamics

The peak in the volume of self-reports corresponded with the time at which vaccinations were sped up (Figure 1). Moreover, after 3 months of vaccinations (the end of February 2021), the popularity of self-reporting started to decrease despite the increasing vaccination roll-out. However, the Spearman correlation coefficient between the volume of self-reports and doses administered from December 9, 2020 until February 28, 2021 was very high (r=0.75, P<.001), and the subsequent count of administered doses increased, while reports on AEs decreased (Figure 1).

Figure 1. Daily counts of reports of adverse events (AE) and doses administered in Russia (data according to Our World in Data [4,44]).



Revealed AE Frequencies (BERT Classes)

Our analysis revealed that fever and generalized pain were the most commonly reported AEs (Table 2). Injection site irritations (local reactions) were an order of magnitude less likely to be reported than fever and pain (systemic reaction). Gastric symptoms (especially diarrhea, with a frequency of 0.6% per report) were less likely to be reported than the average prevalence among the general population (1%-5% for diarrhea [45,46]).

Table 2. Frequencies of mild adverse events extracted from the Telegram group (n=11,515).

Adverse events	n (%)
Systemic	
Fever	5461 (47.43)
Pain	5363 (46.57)
Fatigue	3862 (33.54)
Headache	2855 (24.79)
Chills	2651 (23.02)
Insomnia	600 (5.21)
Lymph node enlargement	186 (1.62)
Local	
Erythema/redness	319 (2.77)
Swelling	206 (1.79)
Pruritis	199 (1.73)
Gastric	
Nausea/vomiting	351 (3.05)
Diarrhea	66 (0.57)

Variations Across Age, Gender, and Dose

Gender was reported by 3992 women and 2762 men. On average, women reported 2.5 AEs (σ=1.79; Q1=1; Q2=2; Q3=4), and men reported 2.1 AEs (σ=1.64; Q1=1; Q2=2; Q3=3). Women reported statistically significantly more AEs (P<.001) according to the results of a Mann-Whitney U test (Table 3).

Age was provided by 6754 users. A linear regression analysis was performed for those who reported being at least 18 years old (minimal age of Russian registration [1]). We found a clear and significant linear relationship (β=.0457, SE=.0014), showing that with every year of life, users reported .0457 fewer AEs (Figure 2). In univariate regression analysis, β is an estimated coefficient with a given SE. Mild AEs among the elderly are known to be less frequently observed for most anti-COVID-19 vaccines [35-37,47].

AEs in response to other anti-COVID-19 vaccines have been found to depend on whether the vaccination was the first or the second dose (if applicable). For instance, AEs in response to mRNA vaccines have tended to be stronger with the second dose [36,37,47]. In contrast, AEs in response to vector vaccines have tended to be milder with the second dose [48,49]. Regarding the Sputnik V vaccine, this difference might be because a different vector is used in each dose, which might lead to different reactions. Among the self-reports, 4174 described AEs after the first dose, 1251 described AEs after the second dose, and 3049 described AEs after both doses. It is also possible that the users did not receive the second dose because of contraindications or just lost interest in reporting.

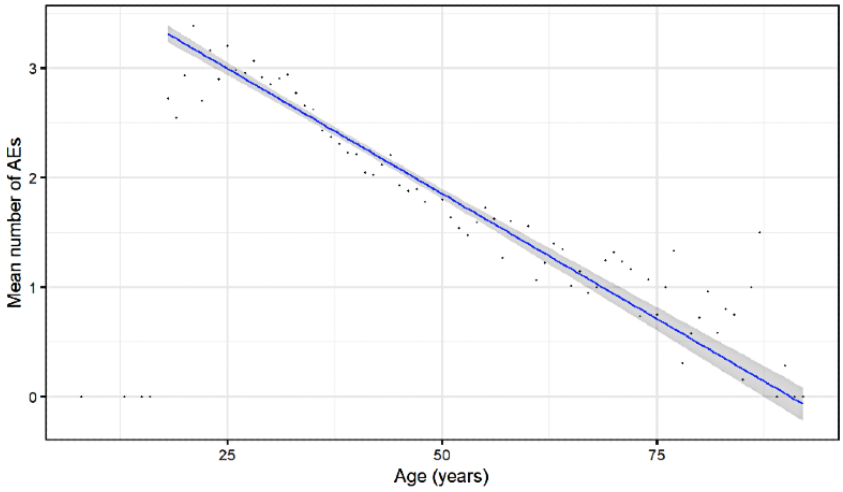
Here, we considered only reports that discussed the first and second doses separately. On average, there were 2.2 (σ=1.80; Q1=0; Q2=2; Q3=4) AEs for the first dose and 1.9 (σ=1.69; Q1=0; Q2=2; Q3=3) AEs for the second dose. According to the results of the Mann-Whitney U test, there were statistically significantly more AEs after the first dose (P<.001; Table 3).

Table 3. Comparisons of the mean numbers of adverse events (AEs) by gender and by dose using Mann-Whitney U tests.

Variable	Number of AEs, mean	OR[a]	*P* value
Gender			
Male	2.1	1.20	<.001
Female	2.5		
Dose			
First	2.2	1.13	<.001
Second	1.9		

[a]OR: odds ratio.

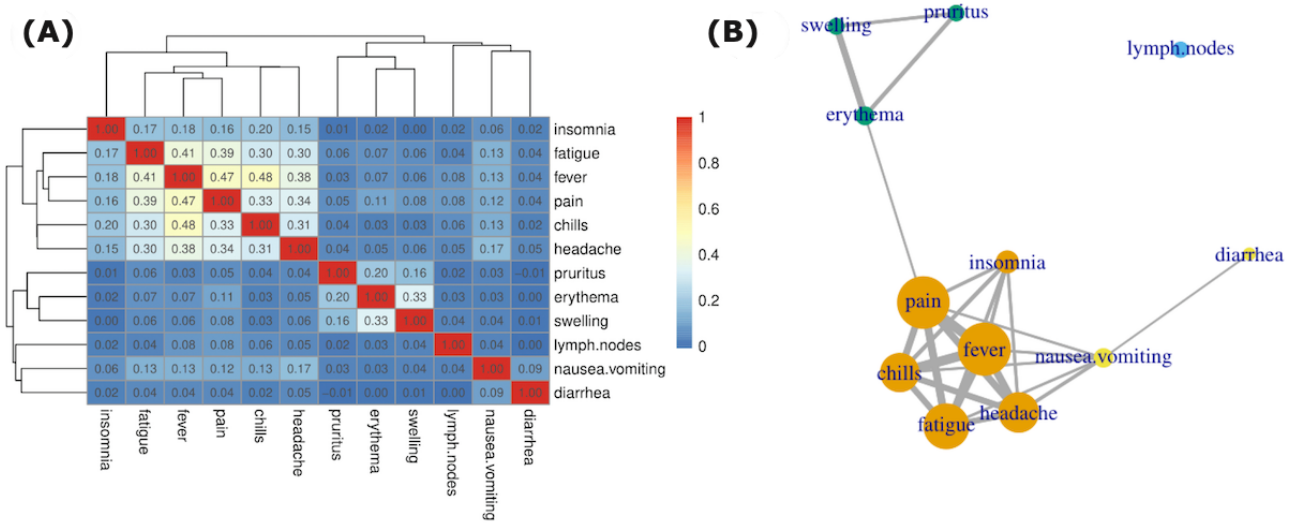
Figure 2. Scatterplot of the number of adverse events (AEs) reported by user vs. age. Dots indicate the mean number of AEs for a given age, while the blue line indicates the linear regression trend and shadowed area indicate its CIs.



Co-occurrence of AEs

To quantify the co-occurrence of symptoms, we calculated Spearman rank correlation coefficients between each pair of classified symptoms. We observed systemic, local, and gastric clusters (Figure 3). We also provided a network representation in which vertex size represents symptom prevalence and edge width represents co-occurrence as measured by the correlation coefficient. Only edges with a correlation coefficient above 0.09 are shown (Figure 3). An unsupervised weighted Louvain algorithm [50] for community detection was used for this purpose, and the vertices were colored the same if they belonged to the same community, which revealed a meaningful structure in which orange denoted systemic, green denoted local, and yellow denoted gastric communities of symptoms.

Figure 3. Co-occurrence of adverse events (AEs), shown as (A) hierarchical clustering based on the correlation matrix of AE symptoms and (B) the corresponding network of AE symptoms with different communities denoted by color code.



Telegram Versus Other Trials or Registries of Sputnik V

We compared our results with 2 available datasets of AEs in response to the Sputnik V vaccine. The first one was collected in Moscow. The second one was collected in Argentina.

Moscow Clinical Trial

Mild AEs in 1029 patients older than 60 years in the phase 3 clinical trial [3,51] in Moscow were compared with 690 self-reports by Telegram users older than 60 years (Table 4). Because there were inconsistencies in various definitions of AEs, a simplified classification was provided, and only headache and diarrhea comprised similar symptoms (at least *sensu lato*). We performed the following calculations to compare both datasets. To obtain *fever* according to our definition, we summed the results for pyrexia, fever sensation, and elevated body temperature from the clinical trial. Similarly, to obtain *pain*, we summed the results for myalgia, arthralgia, and local reaction. To obtain *fatigue*, we summed the results for asthenia and malaise. To obtain *nausea*, we summed the results for nausea and dyspepsia. For *erythema*, we chose the results for contact dermatitis.

In all systemic reactions, Telegram users reported AEs significantly more often than measured in the clinical trial (Table 4). In contrast, diarrhea was less likely to be reported than measured in the clinical trial.

Table 4. Comparisons of adverse events with the Sputnik vaccine between the Telegram and Moscow clinical trial [3] datasets (r=0.69, P=.09).

Adverse event	Moscow clinical trial, n (%)	Telegram, n (%)	OR[a]	P value[b]
Pain	67 (6.70)	177 (25.65)	3.82	<.001
Headache	30 (2.92)	89 (12.90)	4.42	<.001
Fatigue	31 (3.01)	141 (20.43)	6.78	<.001
Fever	32 (3.11)	163 (23.62)	7.59	<.001
Nausea	12 (1.17)	9 (1.30)	1.12	.83
Erythema	39 (3.79)	15 (2.17)	0.57	.09
Diarrhea	8 (0.78)	3 (0.43)	0.56	.54

[a]OR: odds ratio for the Moscow clinical trial.

[b]Fisher test results for the comparison between samples.

Argentinian Postregistration AE Registry

Another available dataset on AEs in response to Sputnik V was compiled from the Argentinian registry of passive AE monitoring (Table 5). This registry contains 23,804 events of all kinds of AEs (mild AEs: 22,971/23,804, 96.5%) from 2,541,362 doses administered. To compare, we chose 7797 Telegram posts that reported at least one AE, and we adjusted new disjoint subsets of symptoms according to the Argentinian methodology [44].

We categorized gastric as the frequency of the logical function nausea OR diarrhea. We categorized site irritation as the frequency of the logical function pruritus OR erythema OR swelling. We categorized fever_pain as the frequency of the logical function fever AND (pain OR headache). We categorized fatigue_pain as the frequency of the logical function fatigue AND (pain OR headache). We categorized only_fever as the frequency of the logical function fever AND ˜(pain OR headache OR fatigue); ˜ denotes logical negation.

The comparison showed that the statistics, despite the significant differences shown in Table 5, were similar in magnitude and highly correlated (r=0.94). The comparison of the Telegram reports (a selected sample with at least one AE constructed by multilabel classification) with the Argentinian registry (multiclass classification [44]) was conducted by the aforementioned mapping. The results of the comparison must be interpreted with caution.

Table 5. Comparisons of adverse events with the Sputnik vaccine between the Telegram and Argentinian safety monitoring [44] datasets (r=0.94, P=.02).

Adverse event	Argentinian registry, n (%)	Telegram, n (%)	OR[a]	P value[b]
fever_pain	8210 (33.25)	4142 (54.70)	1.66	<.001
fatigue_pain	9407 (38.10)	2998 (39.67)	1.05	.05
gastric	1447 (5.98)	395 (5.14)	0.90	.07
site irritation	2306 (9.34)	558 (7.31)	0.80	<.001
only_fever	2065 (8.34)	697 (9.53)	1.11	.02

[a]OR: odds ratio for the Argentinian registry.

[b]Fisher test results for the comparison between samples.

Comparison With Other Vaccines

Regarding vaccines registered by the EMA and FDA, lists of the frequencies of the most common adverse events are accessible; however, they vary across regulatory bodies. Thus, we chose a subset of symptoms for frequencies that were reasonably comparable (pain, headache, fatigue, fever, chills, and nausea). We built a distance (Euclidean) matrix of AEs based on clinical trial registries (EMA [48,52-54], FDA [35-37]) and from the Telegram group (Table 6). From the FDA dataset, for 2-dose vaccines, the dose with higher reactogenicity was selected. In clinical trials, pain is usually considered as pain at the injection site. Fever was the sum of pyrexia and fever in the EMA database. EMA used the injection site tenderness/irritation category. However, regarding redness/erythema, the FDA classified swelling and pruritus separately. Thus, erythema was not included. Sputnik V is a vector vaccine, as are those from AstraZeneca and Johnson & Johnson. The results showed that Telegram Sputnik V AEs were clustered with other vector vaccines, which was possibly due to similar safety profiles (Figure 4).

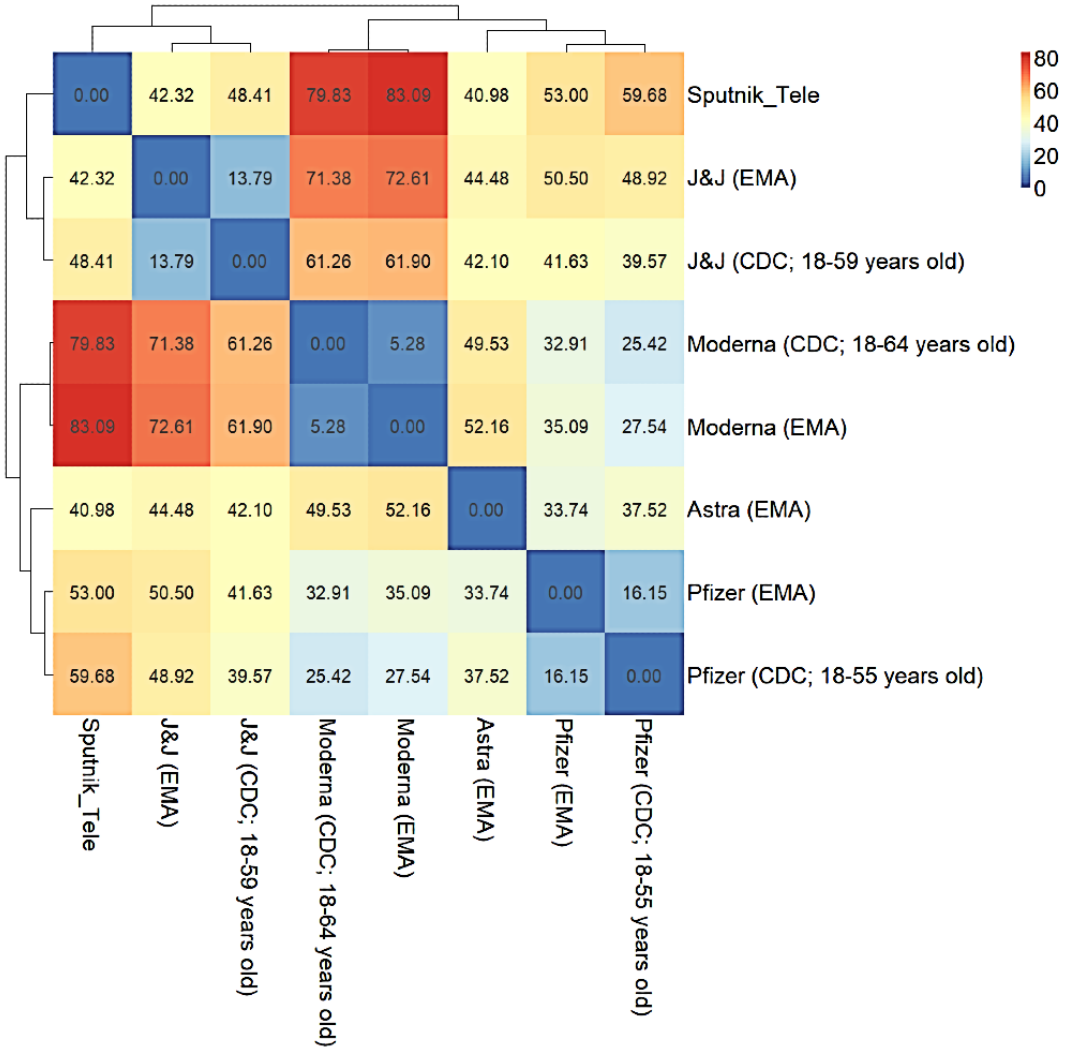
It is important to note that the Telegram users also submitted reports without any AEs at all. Thus, our surveillance system included a sentinel property of samples in contrast to VAERS (North America), EudraVigilance (European Union), and the Argentinian registry [44], which gather reports only if there is any AE to be reported.

Table 6. Adverse events in response to Sputnik V (Telegram) and other vaccines (European Medicines Agency [EMA] and Centers for Disease Control and Prevention [CDC]/Food and Drug Administration [FDA]).

Vaccine	Pain, n (%)	Headache, n (%)	Fatigue, n (%)	Fever, n (%)	Chills, n (%)	Nausea, n (%)
AstraZeneca (EMA)	–[a] (54.20)	– (52.60)	– (53.10)	– (41.50)	– (31.90)	– (21.80)
Johnson & Johnson (EMA)	– (48.60)	– (38.90)	– (38.20)	– (14.00)	– (5.00)	– (14.20)
Johnson & Johnson (CDC; 18-59 years old)	1193 (59.80)	905 (44.40)	891 (43.80)	261 (12.80)	– (5.00)	315 (15.50)
Pfizer (EMA)	– (80.00)	– (50.00)	– (60.00)	– (30.00)	– (30.00)	– (5.00)
Pfizer (CDC; 18-54 years old)	1632 (77.80)	1085 (51.70)	1247 (59.40)	331 (15.80)	737 (35.10)	– (10.00)
Sputnik (Telegram)	5363 (46.57)	2855 (24.80)	3862 (33.54)	5461 (47.43)	2651 (23.02)	351 (3.00)
Moderna (CDC; 18-64 years old)	9335 (90.10)	6500 (62.80)	7002 (67.60)	1806 (17.40)	5001 (48.30)	2209 (21.30)
Moderna (EMA)	– (92.00)	– (64.70)	– (70.00)	– (15.50)	– (45.40)	– (23.00)

[a]Not reported.

Figure 4. Hierarchical matrix of adverse event (AE) similarity of various vaccines and reporting systems (Euclidean distance) of vaccinations investigated in the present study. Astra: AstraZeneca; CDC: Centers for Disease Control and Prevention; EMA: European Medicines Agency; FDA: Food and Drug Administration; J&J: Johnson & Johnson.



Discussion

Principal Findings

According to clinical trials [3] and official registries [44], only partial information could be retrieved on the Sputnik V safety profile. Previously, multiple researchers have raised concerns about the safety of the Sputnik V vaccine [6,55,56]. Our study aimed to increase transparency regarding the safety of Sputnik V [57], because drug regulatory agencies such as in Brazil were delaying Sputnik V emergency registration: “Anvisa was unable to validate the methodology Russian studies used (...) to track

and describe adverse events following vaccination" [58]. In this study, we showed that community-based surveillance via social media can provide meaningful information that could be useful, and this phenomenon should be carefully investigated. The frequencies of AEs extracted from Telegram samples in which at least one AE was reported were in line with other safety surveillance.

Mild, nonsevere AEs have usually been ignored by medical communities because they are common to all vaccines. Antivax movements have emphasized severe AEs, which have been widely discussed in social media [59] in the wider context of vaccine safety [60,61]. In the discourse on COVID-19 vaccines, the main issues were that they were developed quickly and they could compromise safety. Those issues included the fear that vaccines would alter human DNA, cause allergic reactions to vaccine ingredients, result in sudden deaths due to frailty syndrome, or cause infertility [62,63]. Wide anti-COVID-19 immunization programs promulgated a discourse in which risk (eg, the discomfort of common, but mild, AEs as well as rare, but serious AEs) and benefits (eg, efficacy in protecting from the disease) were described as "tradeoffs" of being vaccinated. Mild AEs have become an important issue for many people; moreover, they have the economic component of the potential need for sick leave. This discourse led to the formation of a public Telegram group, where users were asked to report AEs.

In this study, we demonstrated that, in the first phase of the vaccination roll-out, the AE reports were correlated (r=0.7) with vaccination volume (Figure 1). However, Telegram users tended to lose interest after a few months. It is possible that because of the prioritization of vaccine delivery, which began with public and military servants, scientists, teachers, and medical staff, these "early adopters" were more likely to post on social media and be actively involved in reporting AEs. Subsequently, users in the general population were vaccinated, and they were less involved in reporting on the Telegram platform (Figure 1). Thus, interest in COVID-19, Sputnik V, and its AEs was influenced by social context and media to much extent [64].

The results of this study showed that the number of reported AEs decreased linearly according to age (β=.05 AE per year; Figure 2). This result was dependent on biology, which was confirmed in previous clinical trials [36,37,52] and postmarketing observations [47] of other anti-COVID-19 vaccines. Telegram users older than 60 years reported significantly more systemic AEs compared with their peers in clinical trials, who tested negative for or had recovered from COVID-19 [51] (Table 4). On one hand, it is possible that people previously infected with COVID-19 were more likely to report AEs after receiving other vaccines [47]. On the other hand, self-reporting bias could be an important factor in explaining the difference between the "Moscow" clinical trial and the Telegram reports.

The safety profile of the Sputnik V vaccine includes mild AEs that are more similar to vector vaccines than to mRNA anti-COVID-19 vaccines, which was quantified by the Euclidean distance between AE frequencies (Figure 4). The Sputnik V safety profile also showed a high fever-to-fatigue ratio (Table 6) and a stronger reaction to the first dose than to the second one (Table 3), which was also analyzed in a retrospective observational study in San Marino [65].

Women reported more AEs than men (1.2-fold, P<.001; Mann-Whitney U test). This phenomenon is well recorded in other anti-COVID-19 vaccine registries [47,66] and has even been noticed among Argentinian medical staff [67], which could indicate sex-dependent vaccine reactivity. However, this result needs to be understood with caution. The Centers for Disease Control and Prevention has warned that gender bias in reporting could be more important than possible biological mechanisms [35]. The likelihood of disclosing personal information (even anonymously) is known to vary, such as according to gender [68] and social class [69]. A potential reason is that women are more likely to be interested in health, write about health on the internet, and disclose their information [68].

On Telegram, self-reports are most likely to underestimate gastric symptoms (eg, diarrhea at 0.6%). These symptoms could be a taboo effect [70], such as a response to public speaking anxiety. Alternatively, it could be easily ignored because of its high prevalence, or it could be eliminated using an over-the-counter medicine such as loperamide [45,46]. Insomnia was detected so often that it suggests an epidemiological link with the vaccine, which needs further investigation. Local AEs, such as injection site irritation, have rarely been reported. Underlying conditions of erythema/redness, which is usually one of the most common AEs in response to all injected substances including vaccines, are probably overlooked due to low subjective discomfort and lack of physical investigation by a doctor. The findings showed that their actual prevalence was probably underreported.

Our study has several limitations. First, we analyzed participatory and community-based surveillance among Russian Telegram users. Therefore, the results may be specific to the Russian population in a given stage of the pandemic and therefore should not be extrapolated to other contexts. Second, Telegram users may overlook less troublesome side effects, and the social context could influence decisions on taking part in discussions and being selective in reporting AEs [68,69]. For example, local or gastric AEs could be underreported. Third, the classifications developed in this study should not be strictly applied to other contexts. For example, pain at the injection site and pain in other parts of the body were not differentiated. Observed correlations and odds ratios do not imply causation. Fourth, we did not assess the authenticity and credibility of posts [15]; thus, incorrect information could be included in the data. Finally, because our infodemiology study focused on community research initiatives (independent and nonprofit projects, with already known strengths and weaknesses from the history of medicine [71]), our observations cannot replace real-world studies [55-57]. The symptoms reported by social media users only partially reflect their prevalence in the real world [72]. Therefore, the frequencies of symptoms should not be interpreted without considering the contexts and proportions of other symptoms (ie, fever-to-fatigue ratio), phase of the epidemic, and vaccination roll-out (ie, the number of doses administered daily and the population that is vaccinated), as willingness to report AEs satisfies typical product life-cycle temporal characteristics [73,74]).

Conclusion

After the Sputnik V vaccination, Russian Telegram users reported mostly pain, fever, and fatigue. The Sputnik V mild AE profile was comparable with other vector COVID-19 vaccines. Discussions on social media could provide meaningful information about the AE profile of novel vaccines. Further research on severe AEs reported on social media and their credibility is needed.

Acknowledgments

The authors acknowledge the initiators and users of the Telegram group “Sputnik_results” for creation of the data analyzed and express gratitude to the editor and 3 anonymous referees. AJ and VB were partially funded by the German Research Foundation (DFG: 458528774) and Polish-German Foundation for Science (PNFN: 2019-21). Support from The Endowment Fund of St. Petersburg State University is gratefully acknowledged by AS. MK is medicine practitioner at Individual Medical Practice, Oborniki, Poland.

Conflicts of Interest

MK received remuneration for performing vaccinations against COVID-19 in primary care. The vaccinations did not involve Sputnik V.

References

1. Gam-COVID-Vac: Combined vector vaccine for the prevention of coronavirus infection (in Russian). State Register of Medicines. URL: https://grls.rosminzdrav.ru/Grls_View_v2.aspx?routingGuid=6c1f7501-7067-45b3-a56d-95e25db89e97&t [accessed 2021-03-21]
2. Logunov DY, Dolzhikova IV, Zubkova OV, Tukhvatulin A, Shcheblyakov DV, Dzharullaeva AS, et al. Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. The Lancet 2020 Sep 26;396(10255):887-897 [FREE Full text] [doi: 10.1016/S0140-6736(20)31866-3] [Medline: 32896291]
3. Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaeva AS, et al. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. The Lancet 2021 Feb 20;397(10275):671-681 [FREE Full text] [doi: 10.1016/S0140-6736(21)00234-8] [Medline: 33545094]
4. Devlin K, Connaughton A. Most Approve of National Response to COVID-19 in 14 Advanced Economies. Pew Research Center. 2020 Aug 27. URL: https://www.pewresearch.org/global/2020/08/27/most-approve-of-national-response-to-covid-19-in-14-advanced-economies/ [accessed 2021-10-30]
5. The Ministry of Industry and Trade told about the number of vaccines released into circulation in Russia. TASS. 2021 Jun 25. URL: https://tass.ru/ekonomika/11756145 [accessed 2021-10-30]
6. Baraniuk C. Covid-19: What do we know about Sputnik V and other Russian vaccines? BMJ 2021 Mar 19;372:n743. [doi: 10.1136/bmj.n743] [Medline: 33741559]
7. Bucci EM, Berkhof J, Gillibert A, Gopalakrishna G, Calogero RA, Bouter LM, et al. Data discrepancies and substandard reporting of interim data of Sputnik V phase 3 trial. Lancet 2021 May 22;397(10288):1881-1883. [doi: 10.1016/S0140-6736(21)00899-0] [Medline: 33991475]
8. Karafillakis E, Martin S, Simas C, Olsson K, Takacs J, Dada S, et al. Methods for social media monitoring related to vaccination: Systematic scoping review. JMIR Public Health Surveill 2021 Feb 08;7(2):e17149 [FREE Full text] [doi: 10.2196/17149] [Medline: 33555267]
9. Samaras L, García-Barriocanal E, Sicilia MA. Syndromic surveillance using web data: a systematic review. Innovation in Health Informatics 2020:39-77 [FREE Full text] [doi: 10.1016/B978-0-12-819043-2.00002-2]
10. Zheluk A, Gillespie JA, Quinn C. Searching for truth: internet search patterns as a method of investigating online responses to a Russian illicit drug policy debate. J Med Internet Res 2012 Dec 13;14(6):e165 [FREE Full text] [doi: 10.2196/jmir.2270] [Medline: 23238600]
11. Sboev AG, Sboeva SG, Gryaznov AV, Evteeva AV, Rybka RB, Silin MS. A neural network algorithm for extracting pharmacological information from russian-language internet reviews on drugs. J. Phys.: Conf. Ser 2020 Dec 16;1686:012037 [FREE Full text] [doi: 10.1088/1742-6596/1686/1/012037]
12. Dai X, Karimi S, Paris C. Medication and adverse event extraction from noisy text. Proceedings of Australasian Language Technology Association Workshop 2017:79-87 [FREE Full text]
13. Salehan M, Kim D. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems 2016 Jan;81:30-40 [FREE Full text] [doi: 10.1016/j.dss.2015.10.006]
14. Gattepaille LM, Hedfors Vidlin S, Bergvall T, Pierce CE, Ellenius J. Prospective evaluation of adverse event recognition systems in Twitter: Results from the Web-RADR Project. Drug Saf 2020 Aug;43(8):797-808 [FREE Full text] [doi: 10.1007/s40264-020-00942-3] [Medline: 32410156]

15. Hoang T, Liu J, Pratt N, Zheng VW, Chang KC, Roughead E, et al. Authenticity and credibility aware detection of adverse drug events from social media. Int J Med Inform 2018 Dec;120:101-115. [doi: 10.1016/j.ijmedinf.2018.09.002] [Medline: 30409335]
16. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015;1(2):e7 [FREE Full text] [doi: 10.2196/publichealth.4488] [Medline: 27227141]
17. Zhou Z, Hultgren KE. Complementing the US Food and Drug Administration Adverse Event Reporting System with adverse drug reaction reporting from social media: Comparative analysis. JMIR Public Health Surveill 2020 Sep 30;6(3):e19266 [FREE Full text] [doi: 10.2196/19266] [Medline: 32996889]
18. Li Y, Jimeno Yepes A, Xiao C. Combining social media and FDA Adverse Event Reporting System to detect adverse drug reactions. Drug Saf 2020 Sep;43(9):893-903 [FREE Full text] [doi: 10.1007/s40264-020-00943-2] [Medline: 32385840]
19. Patel R, Belousov M, Jani M, Dasgupta N, Winokur C, Nenadic G, et al. Frequent discussion of insomnia and weight gain with glucocorticoid therapy: An analysis of Twitter posts. NPJ Digit Med 2018 Mar 12;1:1 [FREE Full text] [doi: 10.1038/s41746-017-0007-z] [Medline: 30740536]
20. Martínez-López De Castro N, Samartín-Ucha M, Martín-Vila A, Álvarez-Payero M, Piñeiro-Corrales G, Pego-Reigosa JM. Content analysis of Twitter in relation to biological treatments for chronic inflammatory arthropathies: an exploratory study. Eur J Hosp Pharm 2019 May;26(3):124-128 [FREE Full text] [doi: 10.1136/ejhpharm-2017-001402] [Medline: 31428318]
21. Smith K, Golder S, Sarker A, Loke Y, O'Connor K, Gonzalez-Hernandez G. Methods to compare adverse events in Twitter to FAERS, drug information databases, and systematic reviews: Proof of concept with adalimumab. Drug Saf 2018 Dec;41(12):1397-1410 [FREE Full text] [doi: 10.1007/s40264-018-0707-6] [Medline: 30167992]
22. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: An analysis of recent FDA Safety Alerts. Drug Saf 2017 Apr;40(4):317-331 [FREE Full text] [doi: 10.1007/s40264-016-0491-0] [Medline: 28044249]
23. Huesch MD. Commercial online social network data and statin side-effect surveillance: A pilot observational study of aggregate mentions on Facebook. Drug Saf 2017 Dec;40(12):1199-1204. [doi: 10.1007/s40264-017-0577-3] [Medline: 28748367]
24. Telegram audience grew by 1.2 million people in a month. TASS. 2020 Oct 16. URL: https://tass.ru/obschestvo/9743857 [accessed 2021-10-30]
25. Melkadze A. Leading social media platforms in Russia as of 3rd quarter of 2020, by penetration rate. Statista. 2021 Mar 18. URL: https://www.statista.com/statistics/867549/top-active-social-media-platforms-in-russia/ [accessed 2021-10-30]
26. Wang J, Zhao L, Ye Y, Zhang Y. Adverse event detection by integrating twitter data and VAERS. J Biomed Semantics 2018 Jun 20;9(1):19 [FREE Full text] [doi: 10.1186/s13326-018-0184-y] [Medline: 29925405]
27. Jarynowski A. Sputnik V Adverse Events risk calculator. 2021. URL: https://infodemia-koronawirusa.shinyapps.io/sputnik/ [accessed 2021-06-05]
28. People's reports on vaccination from Covid-19: Project V1V2.ru | reviews after vaccine vaccination against covid side effects satellite v m light. Telegram. URL: https://t.me/Sputnik_results [accessed 2021-10-30]
29. Terms of Service. Telegram. URL: https://telegram.org/tos [accessed 2021-10-30]
30. Semenov A, Mantzaris A, Nikolaev A, Veremyev A, Veijalainen J, Pasiliao E, et al. Exploring social media network landscape of post-Soviet space. IEEE Access 2019;7:411-426 [FREE Full text] [doi: 10.1109/access.2018.2885479]
31. LonamiWebs / Telethon. GitHub. URL: https://github.com/LonamiWebs/Telethon [accessed 2021-03-05]
32. EudraVigilance - European database of suspected adverse drug reaction reports. European Medicines Agency. URL: https://www.adrreports.eu/ [accessed 2021-10-30]
33. Vaccine Adverse Event Reporting System. Department of Health and Human Services. URL: https://vaers.hhs.gov/ [accessed 2021-10-30]
34. Yellow Card. Medicines and Healthcare products Regulatory Agency. URL: https://yellowcard.mhra.gov.uk/ [accessed 2021-10-30]
35. The Janssen COVID-19 Vaccine's Local Reactions, Systemic Reactions, Adverse Events, and Serious Adverse Events. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/covid-19/info-by-product/janssen/reactogenicity.html [accessed 2021-05-01]
36. The Moderna COVID-19 Vaccine's Local Reactions, Systemic Reactions, Adverse Events, and Serious Adverse Events. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/covid-19/info-by-product/moderna/reactogenicity.html [accessed 2021-03-21]
37. The Pfizer COVID-19 Vaccine's Local Reactions, Systemic Reactions, Adverse Events, and Serious Adverse Events. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/covid-19/info-by-product/pfizer/reactogenicity.html [accessed 2021-03-21]
38. McDonald I, Murray SM, Reynolds CJ, Altmann DM, Boyton RJ. Comparative systematic review and meta-analysis of reactogenicity, immunogenicity and efficacy of vaccines against SARS-CoV-2. NPJ Vaccines 2021 May 13;6(1):74 [FREE Full text] [doi: 10.1038/s41541-021-00336-1] [Medline: 33986272]

39. Kaur RJ, Dutta S, Bhardwaj P, Charan J, Dhingra S, Mitra P, et al. Adverse events reported from COVID-19 vaccine trials: A systematic review. Indian J Clin Biochem 2021 Mar 27:1-13 [FREE Full text] [doi: 10.1007/s12291-021-00968-z] [Medline: 33814753]
40. Label Studio. Heartex. URL: https://labelstud.io/ [accessed 2021-10-30]
41. Devlan J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Cornell University. 2019 May 24. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-10-30]
42. deepmipt / DeepPavlov. GitHub. URL: https://github.com/deepmipt/DeepPavlov [accessed 2021-04-05]
43. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management 2009 Jul;45(4):427-437 [FREE Full text] [doi: 10.1016/j.ipm.2009.03.002]
44. 10th Vaccine Safety Report. Ministerio de Salud Argentina. URL: https://bancos.salud.gob.ar/recurso/10deg-informe-de-seguridad-en-vacunas [accessed 2021-05-01]
45. Kamiński M, Borger M, Prymas P, Muth A, Stachowski A, Łoniewski I, et al. Analysis of answers to queries among anonymous users with gastroenterological problems on an internet forum. Int J Environ Res Public Health 2020 Feb 06;17(3):1 [FREE Full text] [doi: 10.3390/ijerph17031042] [Medline: 32041356]
46. Panayiotou G, Karekla M, Georgiou D, Constantinou E, Paraskeva-Siamata M. Psychophysiological and self-reported reactivity associated with social anxiety and public speaking fear symptoms: Effects of fear versus distress. Psychiatry Res 2017 Sep;255:278-286. [doi: 10.1016/j.psychres.2017.05.044] [Medline: 28599192]
47. Menni C, Klaser K, May A, Polidori L, Capdevila J, Louca P, et al. Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study. Lancet Infect Dis 2021 Jul;21(7):939-949 [FREE Full text] [doi: 10.1016/S1473-3099(21)00224-3] [Medline: 33930320]
48. COVID-19 Vaccine AstraZeneca - Summary of Product Characteristics. European Medicines Agency. URL: https://ec.europa.eu/health/documents/community-register/2021/20210129150842/anx_150842_en.pdf [accessed 2021-03-21]
49. Coronavirus vaccine - weekly summary of Yellow Card reporting. Medicines & Healthcare products Regulatory Agency. URL: https://www.gov.uk/government/publications/coronavirus-covid-19-vaccine-adverse-reactions/coronavirus-vaccine-summary-of-yellow-card-reporting [accessed 2021-09-02]
50. Blondel V, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J. Stat. Mech 2008 Oct 09;2008(10):P10008 [FREE Full text] [doi: 10.1088/1742-5468/2008/10/p10008]
51. Logunov DY, Dolzhikova IV, Tukhvatullin AI, Shcheblyakov DV. Safety and efficacy of the Russian COVID-19 vaccine: more information needed - Authors' reply. Lancet 2020 Oct 03;396(10256):e54-e55 [FREE Full text] [doi: 10.1016/S0140-6736(20)31970-X] [Medline: 32971043]
52. COVID-19 Vaccine Janssen-Summary of Product Characteristics. European Commission. URL: https://www.ema.europa.eu/en/documents/product-information/covid-19-vaccine-janssen-epar-product-information_en.pdf [accessed 2021-05-01]
53. COVID-19 mRNA Vaccine Moderna-Summary of Product Characteristics. European Commission. URL: https://ec.europa.eu/health/documents/community-register/2021/20210106150575/anx_150575_en.pdf [accessed 2021-03-21]
54. COVID-19 mRNA Vaccine Comirnaty-Summary of Product Characteristics. European Medicines Agency. URL: https://www.ema.europa.eu/en/documents/product-information/comirnaty-epar-product-information_en.pdf [accessed 2021-03-21]
55. van Tulleken C. Covid-19: Sputnik vaccine rockets, thanks to Lancet boost. BMJ 2021 May 06;373:n1108. [doi: 10.1136/bmj.n1108] [Medline: 33958384]
56. Vlassov V. Sputnik V and Russia's covid-19 vaccine race. The BMJ Opinion. 2021 May 06. URL: https://blogs.bmj.com/bmj/2021/05/06/vasiliy-vlassov-sputnik-v-and-russias-covid-19-vaccine-race [accessed 2021-05-01]
57. Logunov DY, Dolzhikova IV, Shcheblyakov DV. Data discrepancies and substandard reporting of interim data of Sputnik V phase 3 trial - Authors' reply. Lancet 2021 May 22;397(10288):1883-1884 [FREE Full text] [doi: 10.1016/S0140-6736(21)00894-1] [Medline: 33991474]
58. Brito R, Ivanova P. Brazil health regulator rejects Russia's Sputnik vaccine. Reuters. 2021 Apr 26. URL: https://www.reuters.com/world/americas/brazil-health-regulator-anvisas-technical-staff-recommend-against-importing-2021-04-26/ [accessed 2021-05-01]
59. Klimiuk K, Czoska A, Biernacka K, Balwicki. Vaccine misinformation on social media - topic-based content and sentiment analysis of Polish vaccine-deniers' comments on Facebook. Hum Vaccin Immunother 2021 Jul 03;17(7):2026-2035 [FREE Full text] [doi: 10.1080/21645515.2020.1850072] [Medline: 33517844]
60. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
61. Cianciara D, Szmigiel A. Posting on „Nie szczepimy („We don't vaccinate") internet forum. Przegl Epidemiol 2019;73(1):105-115 [FREE Full text] [doi: 10.32394/pe.73.11] [Medline: 31134779]
62. Griffith J, Marani H, Monkman H. COVID-19 vaccine hesitancy in Canada: Content analysis of tweets using the Theoretical Domains Framework. J Med Internet Res 2021 Apr 13;23(4):e26874 [FREE Full text] [doi: 10.2196/26874] [Medline: 33769946]
63. Thelwall M, Kousha K, Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. EPI 2021 Mar 15:1 [FREE Full text] [doi: 10.3145/epi.2021.mar.12]

64. Rovetta A. Reliability of Google Trends: Analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. Front Res Metr Anal 2021;6:670226 [FREE Full text] [doi: 10.3389/frma.2021.670226] [Medline: 34113751]
65. Montalti M, Soldà G, Di Valerio Z, Salussolia A, Lenzi J, Forcellini M, San Marino Republic COVID ROCCA Group. ROCCA observational study: Early results on safety of Sputnik V vaccine (Gam-COVID-Vac) in the Republic of San Marino using active surveillance. EClinicalMedicine 2021 Aug;38:101027 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101027] [Medline: 34505029]
66. Gee J, Marquez P, Su J, Calvert GM, Liu R, Myers T, et al. First month of COVID-19 vaccine safety monitoring - United States, December 14, 2020-January 13, 2021. MMWR Morb Mortal Wkly Rep 2021 Mar 26;70(8):283-288 [FREE Full text] [doi: 10.15585/mmwr.mm7008e3] [Medline: 33630816]
67. Pagotto V, Ferloni A, Mercedes Soriano M, Díaz M, Braguinsky Golde N, González MI, et al. Active monitoring of early safety of Sputnik V vaccine in Buenos Aires, Argentina. Medicina (B Aires) 2021;81(3):408-414 [FREE Full text] [Medline: 34137701]
68. Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Søndergaard J, Rasmussen S, et al. Self-reported symptoms and healthcare seeking in the general population--exploring "The Symptom Iceberg". BMC Public Health 2015 Jul 21;15:685 [FREE Full text] [doi: 10.1186/s12889-015-2034-5] [Medline: 26195232]
69. Sæbø O, Federici T, Braccini AM. Combining social media affordances for organising collective action. Information Systems Journal 2020;30(4):699-732 [FREE Full text] [doi: 10.1111/isj.12280]
70. Kowalski P. About the inevitabl excrement and defecation (in Polish). Colloquia Anthropologica et Communicativa: Ciało cielesne 2011:19-41 [FREE Full text]
71. Epstein S. Impure science: AIDS, activism, and the politics of knowledge. Med Soc (Berkeley) 1996:1-466. [Medline: 11619509]
72. Shimabukuro TT, Kim SY, Myers TR, Moro PL, Oduyebo T, Panagiotakopoulos L, CDC v-safe COVID-19 Pregnancy Registry Team. Preliminary findings of mRNA Covid-19 vaccine safety in pregnant persons. N Engl J Med 2021 Jun 17;384(24):2273-2282 [FREE Full text] [doi: 10.1056/NEJMoa2104983] [Medline: 33882218]
73. Rogers EM. Diffusion of innovations. New York City, NY: Simon and Schuster; 2010.
74. Gopalsamy R, Semenov A, Pasiliao E, McIntosh S, Nikolaev A. Engagement as a driver of growth of online health forums: Observational study. J Med Internet Res 2017 Aug 29;19(8):e304 [FREE Full text] [doi: 10.2196/jmir.7249] [Medline: 28851677]

Abbreviations

AE: adverse event
ANN: artificial neural network
AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
EMA: European Medicines Agency
FDA: Food and Drug Administration
LSTM: long short-term memory
MHRA: Medicines and Healthcare products Regulatory Agency
ROC: receiver operating characteristic
VAERS: Vaccine Adverse Event Reporting System

Edited by G Eysenbach; submitted 18.05.21; peer-reviewed by S Wei, B Vasavada, A Rovetta; comments to author 21.06.21; revised version received 12.08.21; accepted 28.09.21; published 29.11.21.

Please cite as:
Jarynowski A, Semenov A, Kamiński M, Belik V
Mild Adverse Events of Sputnik V Vaccine in Russia: Social Media Content Analysis of Telegram via Deep Learning
J Med Internet Res 2021;23(11):e30529
URL: https://www.jmir.org/2021/11/e30529
doi:10.2196/30529
PMID:34662291

Review

Telemedicine in Intensive Care Units: Scoping Review

Camille Guinemer[1], MA; Martin Boeker[2], MSc, MD; Daniel Fürstenau[1,3], PhD; Akira-Sebastian Poncette[1], MD; Björn Weiss[4], MD; Rudolf Mörgeli[4], BSc, MD; Felix Balzer[1], MSc, MD, PhD

[1]Institute of Medical Informatics, Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Intelligence and Informatics in Medicine, Medical Center rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

[3]Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark

[4]Department of Anesthesiology and Operative Intensive Care Medicine (CCM, CVK), Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

Corresponding Author:
Camille Guinemer, MA
Institute of Medical Informatics
Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin
Humboldt-Universität zu Berlin, and Berlin Institute of Health
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 57042
Email: camille.guinemer@charite.de

Abstract

Background: The role of telemedicine in intensive care has been increasing steadily. Tele–intensive care unit (ICU) interventions are varied and can be used in different levels of treatment, often with direct implications for the intensive care processes. Although a substantial body of primary and secondary literature has been published on the topic, there is a need for broadening the understanding of the organizational factors influencing the effectiveness of telemedical interventions in the ICU.

Objective: This scoping review aims to provide a map of existing evidence on tele-ICU interventions, focusing on the analysis of the implementation context and identifying areas for further technological research.

Methods: A research protocol outlining the method has been published in JMIR Research Protocols. This review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews). A core research team was assembled to provide feedback and discuss findings.

Results: A total of 3019 results were retrieved. After screening, 25 studies were included in the final analysis. We were able to characterize the context of tele-ICU studies and identify three use cases for tele-ICU interventions. The first use case is *extending coverage*, which describes interventions aimed at extending the availability of intensive care capabilities. The second use case is *improving compliance*, which includes interventions targeted at improving patient safety, intensive care best practices, and quality of care. The third use case, *facilitating transfer*, describes telemedicine interventions targeted toward the management of patient transfers to or from the ICU.

Conclusions: The benefits of tele-ICU interventions have been well documented for centralized systems aimed at extending critical care capabilities in a community setting and improving care compliance in tertiary hospitals. No strong evidence has been found on the reduction of patient transfers following tele-ICU intervention.

International Registered Report Identifier (IRRID): RR2-10.2196/19695

(J Med Internet Res 2021;23(11):e32264) doi:10.2196/32264

KEYWORDS

tele-ICU; telemedicine; critical care; implementation; telehealth; health care system; intensive care unit; health technology; digital health; care compliance; tertiary hospitals; hospital; review

Introduction

Telemedicine has been increasingly used in intensive care, and approximately 15% of intensive care beds in the United States currently partake in telemedical programs [1-3]. A range of rationales for the implementation of telemedical systems in intensive care has been suggested. Tele–intensive care unit (ICU) technologies have been used to address staffing shortage in intensive care and as a cost-effective response not only to a lack of intensive care availability in some areas but also as a means of increasing adherence to evidence-based best practices using benchmark performance data [3-5].

The American Telemedicine Association defines tele-ICU as "a network of audiovisual communication and computer systems that provide the foundation for a collaborative, interprofessional care model focusing on critically ill patients" [3]. Tele-ICU interventions are varied, can be offered in different levels of intensive care service, and can be customized to meet the specific intensive care needs of hospitals [3,5-7]. For example, some tele-ICU systems provide 24/7 remote monitoring staffed by intensivists, while other systems provide scheduled remote intensivist consultations during nighttime only.

The main characteristics of tele-ICU systems have been well described in the literature. First, technical architectures can be described as centralized or decentralized. Centralized architecture features a command center, or a *cockpit*, connecting one or multiple centers. Decentralized systems (also named *virtual consultant*) allow one-on-one connections without the need for central coordination [3]. Second, staff allocation and availability can vary (eg, day presence or 24/7) [8]. Third, the mode of interaction between telemedicine teams and bedside staff may allow various levels of staff reactivity (reactive vs proactive to patient alerts) and intervention scope (minimal intervention allowed vs full discretion on patient care) [4]. Several guidelines, such as the US [3] or the German Guidelines for Telemedicine in Intensive Medicine [9], provide general recommendations on aspects of equipment, staffing, and organization for implementing tele-ICU systems.

A significant body of primary and secondary literature has been published on ICU telemedical interventions [10]. To date, 9 systematic reviews and 9 other review types have been published on this topic [11], as well as 3 meta-analyses with a focus on medical outcomes (eg, hospital mortality and length of stay) [12]. In previous reviews, the results of tele-ICU interventions have been characterized as heterogeneous [13,14]. Although positive medical outcomes could be detected in some interventions, other contexts could only demonstrate mixed or no positive results at all [4,14,15]. Authors have suggested that the context of implementation may be a factor in explaining the variability of these results. We define context of implementation as the clinical structures and processes where telemedical interventions are deployed [16]. It has been suggested that the efficacy of tele-ICU interventions is dependent on where and how they are deployed in the organization [6,10], and there is a need for broadening the understanding of the organizational factors influencing the efficacy of tele-ICU interventions [8]. We found that no previous study has attempted to provide a review of current evidence by systematically analyzing the implementation setup and context.

This scoping review seeks to address a research gap on the characterization of the context of implementation for tele-ICU interventions [14,17]. The first objective is to characterize the implementation context of tele-ICU interventions with a consistent set of domains on hospital organization. The second objective is to characterize the configurations and structures of tele-ICU systems in relation to their context of implementation. The third objective is to describe the outcomes of tele-ICU interventions and to characterize current evidence according to their intervention contexts.

Methods

A research protocol for this review was published in JMIR Research Protocols in December 2020 [11], which was developed in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) and best practices advanced by Arksey and O'Malley [18] and the Joanna Briggs Institute [19]. The method included the steps *identification of relevant studies*, *selection of study*, *data charting*, and *data collating*.

For the step *identification of relevant studies*, a search for peer-reviewed studies in the databases Web of Science Core Collection, MEDLINE, ERIC, PsycINFO, PSYNDEX, CINAHL, and IEEE was performed without date restrictions. Manual searches were performed additionally to identify gray literature. The search query was developed according to the guidelines of the Peer Review of Electronic Search Strategies and included keywords on the topics of intensive care and telemedicine. The full queries are provided in Multimedia Appendix 1. The search records were downloaded in the reference software Citavi version 6 (Swiss Academic Software).

In the step *selection of study*, both titles and abstracts were screened, and studies not dealing with a relevant topic or method were removed. Results were then screened to find articles where the PICO (Patient, Intervention, Comparison, Outcome) framework could be identified. We included articles with at least three of the PICO criteria summarized in Textbox 1. Studies concerning interventions in neonatal and pediatric ICUs were excluded from this scoping review.

In the step *data charting*, article information was collected and classified into extraction sheets according to the five domains defined in the review protocol (see Textbox 2).

In the step *data collating, summarizing, and reporting*, the information was organized and clustered into an evidence map. The evidence map provided a summary of the scoping review results. During the review process, a core research team was created to provide feedback and discuss findings. The research team was composed of a doctoral researcher with a background in health economics (author CG), a medical data science professor (author FB), a medical informatics professor (author MB), an anesthesiologist with intensive care specialty and main coordinator of a tele-ICU project (author BW), an anesthesiology researcher with a specialty in digital health (author ASP), a professor of digitalization (author DF), and an anesthesiologist

with intensive care specialty (author RM). The research team was asked to consider the information from data charting, provide insights, and discuss results. Differing views were resolved through discussion until consensus was reached.

Textbox 1. PICO (Patients, Intervention, Comparison, Outcomes) criteria.

Patient

Participants provided telemedical intensive care.

Intervention

Telemedical system implemented with one more an intensive care units (ICUs).

Comparison

Comparison with the standard of care without tele-ICU intervention.

Outcomes

All outcomes eligible for inclusion.

Textbox 2. Data charting domains.

Implementation context

A. Clinical focus

Level of intensive care specialization. Generalist (medical intensive care unit [ICU], surgical ICU) or specialized clinical focus (ie, sepsis, cardiology, neurocritical).

B. ICU type

Level of intensivist involvement in patient care. Defined by staffing model of ICU (ie, open vs closed ICU models).

C. Hospital type

Category of hospital involved in tele-ICU intervention (ie, tertiary or community hospital). Community hospitals are defined as nonfederal, short-term general hospitals under 500 beds [20].

D. System configuration

Technical architecture (ie, centralized vs decentralized), staff allocation (ie, continuous vs scheduled), and mode of communication of the tele-ICU system (ie, high or low data intensity).

E. Implementation rationale

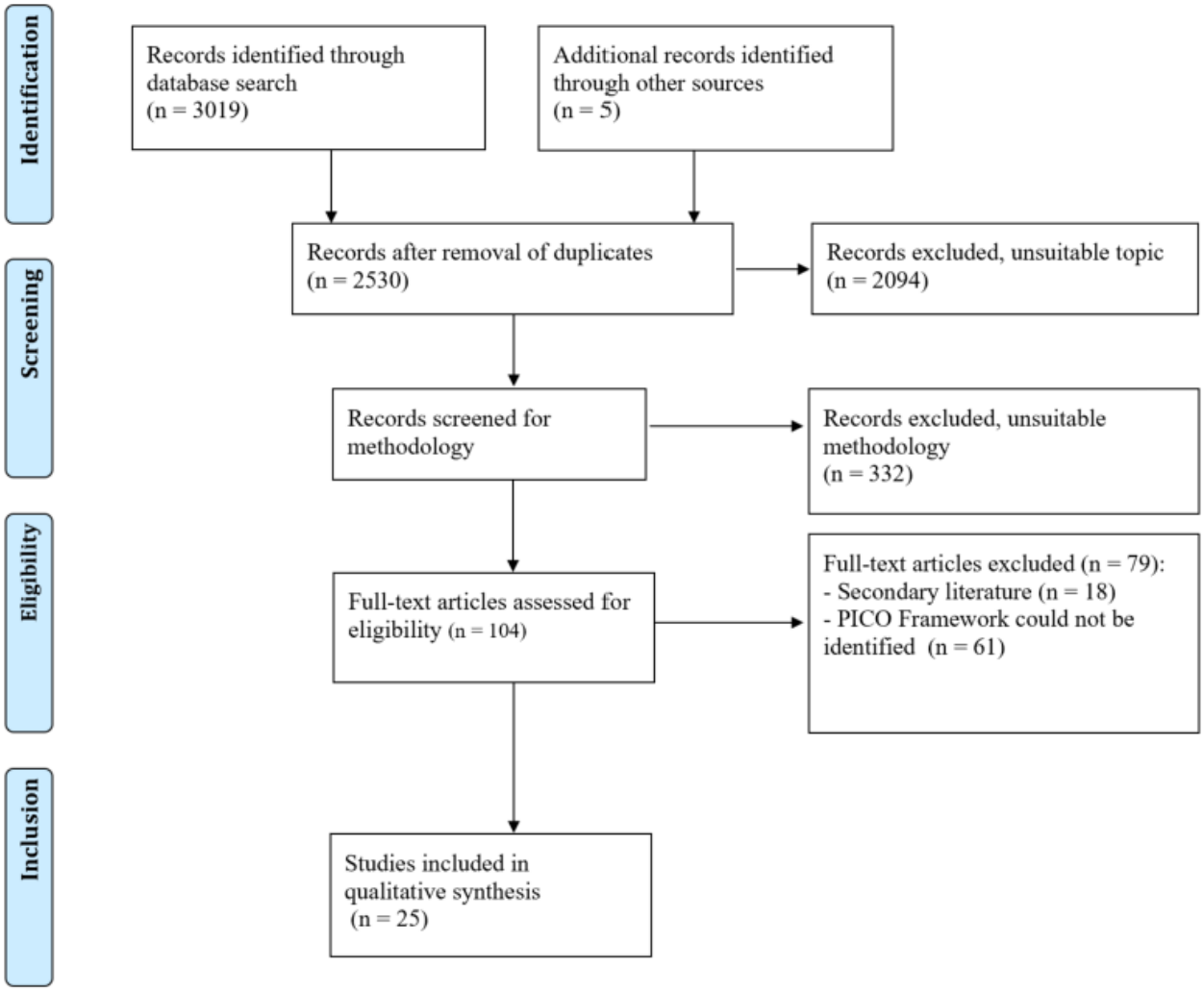
Main rationale provided in the study for tele-ICU intervention, use case for telemedical system in the ICU.

Results

Selection of Relevant Studies

The flowchart in Figure 1 outlines the records yielded by the search. A total of 3019 results were retrieved, including 489 duplicates. After screening, 104 records were eligible for full-text analysis and 25 were included in the final analysis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. PICO: Patient, Intervention, Comparison, Outcomes.



Characteristics of Tele-ICU Studies Included in the Scoping Review

The 25 studies included in this review were published between 2004 and 2019. Out of 25 articles, 21 (84%) referred to tele-ICU implementation within the Unites States, while the remaining papers described implementation in Germany, India, Australia, and Saudi Arabia. Regarding the research methods used in the studies, we found that 21 articles used pre-post comparison designs, of which only 7 included a control group. The pre-post design has been described as a quasi-experimental research design [12,21], for which a random assignment of patients between treatment and control group was not performed. The remaining 4 publications used other methods, such as interrupted time series, and half of these included a control group. We found no examples of randomized controlled trials.

Results From Data Charting

Table 1 summarizes the data charting results for the 5 research domains and provides definitions for each category.

First, we outline results for the domains pertaining to context of implementation (domains A to C). For domain A, most telemedical implementations did not have a specific clinical focus (n=21, 84% of the studies), with only a few cases of specialized interventions. For domain B, tele-ICU interventions were predominantly implemented in ICUs featuring aspects of the open model. In these interventions, the primary physicians or surgeons retained full responsibility for the patient (n=10, 40% of the studies) or with limited intensivist involvement only (n=9, 36% of cases open/closed). Regarding domain C, although 44% (n=11 studies) of interventions were implemented in tertiary hospitals, a large subset was in community settings and in organizations spanning both tertiary and community hospital settings.

Second, concerning the system configuration results in domain D, centralized architectures (eg, tele-ICU Command Center) were the predominant implementation setup. Relating to the staffing model, the continuous care setup was used in 13 (52%) of the studies, where the remote care team assumes constant patient monitoring. Scheduled interventions (eg, daily intensive care rounds) were found in 9 (36%) cases. Finally, most telemedical systems (n=19, 76%) enabled remote real-time access to patient data. To summarize this information, we classified the system configurations into three clusters, as outlined in Figure 2.

Finally, concerning the implementation rationale defined in domain E, three main use cases were defined for tele-ICU interventions. We classified 13 (52%) publications under the use case 1 summarized by the term *extending coverage*. In this group, studies cited intensivist shortage, need for additional intensivist coverage, and extension of intensivist resources as a rationale for the intervention. A total of 10 (40%) studies were classified under use case 2, summarized by the term *improving compliance*. In this group, studies cited the increase in adherence to compliance with care bundles, clinical practice guidelines, or care quality initiatives as the main rationale. We classified two studies in use case 3, summarized by the term *facilitating transfer*. Studies in this category cited the screening or monitoring of patients prior to transfer to or from an ICU as the main rationale.

Table 1. Data charting results: interventions and context.

Domain and category	Definition	Studies (N=25), n (%)
Implementation context		
Clinical focus		
General	No specific clinical focus identified (MICU[a], SICU[b])	21 (84)
Specialized	Specific clinical focus (ie, sepsis, cardiology, neurocritical)	4 (16)
ICU[c] type		
Open	Primary physician has full-time responsibility for patient care	10 (40)
Open/closed	Features of both open and closed models	9 (36)
Closed	Intensivists available with full responsibility for patient care	6 (24)
Hospital type		
Tertiary	Tertiary care institutions or teaching hospitals	11 (44)
Mixed	Care organization spanning tertiary and community settings	4 (16)
Community	Community hospitals or small medical facility	9 (36)
Not available	N/A[d]	1 (4)
System configuration		
Continuous	Continuous patient critical care monitoring	5 (20)
Mixed	Continuous monitoring including scheduled rounds	9 (36)
Scheduled	Scheduled consultation at regular interval. Virtual rounds.	9 (36)
Not available	Insufficient information provided	2 (8)
Centralized	Tele-ICU Command Center or Hub centralizing patient care	19 (76)
Decentralized	Distributed architecture without centralized hub	5 (20)
Not available	N/A	1 (4)
Direct access	Direct staff remote access to patient data	18 (72)
Limited access	Limited staff remote access (screen sharing) to patient data	4 (16)
Not available	N/A	3 (12)
Implementation rationale		
Coverage	Intensivist shortage, provision of extended coverage	13 (52)
Compliance	Adherence and compliance to critical care guidelines	10 (40)
Transfer	Patients screening or triage for transfers to or from ICU	2 (8)

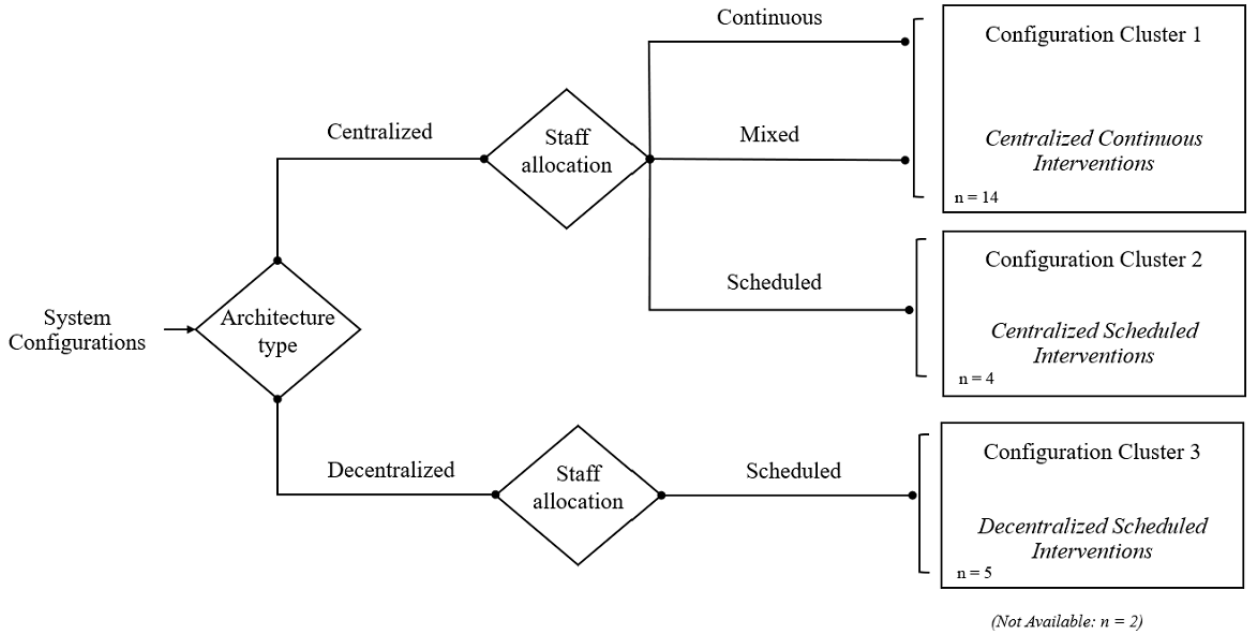
[a]MICU: medical intensive care unit.

[b]SICU: surgical intensive care unit.

[c]ICU: intensive care unit.

[d]N/A: not applicable.

Figure 2. Clustering system configurations.



Intervention Outcomes

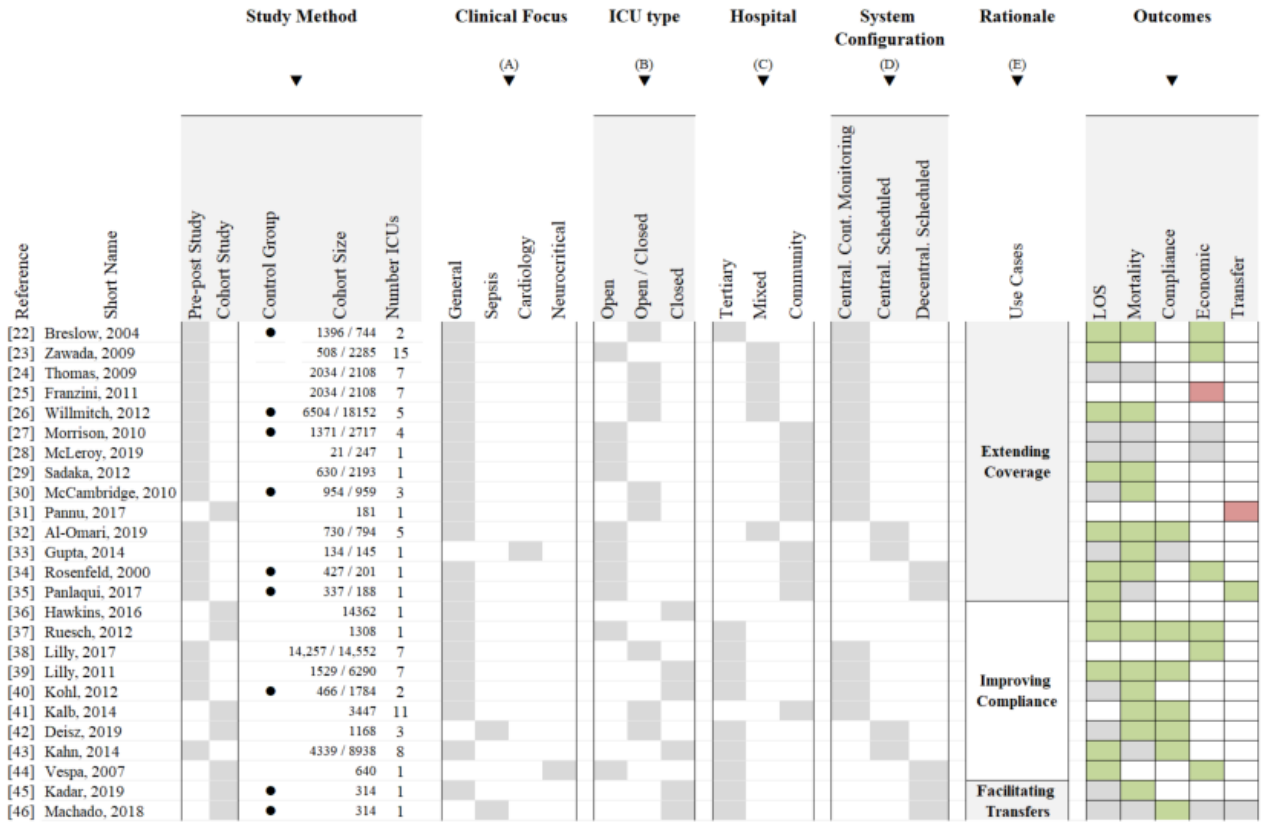
This section presents results on the range of outcomes that were found in the studies on ICU implementation, which are summarized in Table 2.

First, a significant subset of studies provided results on at least one medical outcome. Effect of tele-ICU intervention on length of stay (LOS) was reported in 21 (84%) studies. This outcome was defined as the number of inpatient days for the episode of care in the ICU or in aggregate in the hospital. Results on mortality rates were provided in 19 (76%) studies, including ICU and hospital mortality. In 12 studies, reduction in LOS was found to be significant. Reduction in mortality was significant in 13 studies. Second, 8 (32%) studies measured the rate of adherence to best practices and guidelines implementation, summarized by the term *compliance*. A large subset indicated a statistically significant increase in the level of adherence to ICU standards. Third, under the header *economics*, 9 (36%) studies provided results regarding cost-effectiveness of tele-ICU interventions. In this subset, 6 (67%) studies reported interventions as being cost-effective. Lastly, 2 studies in the category *transfer* measured changes in rate of patient transfer following intervention. One study measured the number of transfers within the same facility (eg, for preadmission diagnostic) and another the number of transfers to another facility (eg, for advanced care). Finally, we note that none of the studies included patient satisfaction scores. These results are summarized as an evidence map in Figure 3.

Table 2. Data charting results: outcomes.

Outcome category	Reporting on outcome, n	Of which reporting positive results, n
Length of stay	21	12
Mortality	19	13
Compliance	8	7
Economics	9	6
Transfer	2	1

Figure 3. Evidence map [22-46]. ICU: intensive care unit; LOS: length of stay.



Discussion

Principal Findings

This scoping review provided an overview of the literature on telemedical interventions in the ICU. Based on a set of defined domains, we were able to characterize the context of tele-ICU studies and identify three use cases for tele-ICU interventions. This analysis aimed to identify common features within the heterogenous use of telemedical systems. Recent research findings relevant for implementation under each use case were outlined.

The first use case, *extending coverage*, included interventions aimed at increasing intensive care coverage in contexts where it is not (or only partially) available at the bedside. This use case was found predominantly in community hospitals having limited onsite critical care capacity. The second use case, *improving compliance*, included interventions targeted at improving patient safety, intensive care best practices and quality of care. These interventions were found primarily in tertiary care context. The third use case, *facilitating transfer*, included telemedicine interventions targeting toward the management of patient transfers to or from the ICU.

Use Case: Extending Coverage

Interventions were predominantly found in community hospitals and in mixed community/tertiary contexts (eg, hospital groups spanning one or several community branches). Tele-ICU systems in this use case have been used to address specific issues related to the delivery of critical care in community and rural areas. Particularly in the United States, recent surveys indicate that hospitalists (ie, physicians whose main focus is on general medical care of patients who are hospitalized [47]) are still the main physician in rural and community settings, reflecting a general shortage in intensive care staffing [48]. Although community hospitals face difficulties in hiring qualified critical care personnel, some of them are subject to minimum requirements to have full intensivist staffing during the day [49,50]. In underserved areas, tele-ICU implementation can therefore represent a valuable solution for the onsite provision of intensive care expertise [51].

The predominant tele-ICU system configuration in this use case was a centralized system featuring continuous remote staff intervention from a workstation, with direct involvement in patient care (configuration cluster: *centralized continuous monitoring*). Team cooperation and sharing of responsibility over patient care between the bedside and remote team are central issues in this type of configuration. Our analysis showed that different modalities of a remote care team have been implemented. In some interventions, the main role of the remote team was to consult and advise the bedside team (Zawada et al [23], McLeroy et al [28], and Al-Omari et al [32]; n=3, 12% of studies), whereas in other cases, remote staff were granted a different level of authority on patient care at the discretion of the bedside team (Sadaka et al [29], Morrison et al [27], Thomas [24], Willmitch et al [26], Franzini et al [25], and Breslow et al [22]; n=6, 24% of studies). Achieving an appropriate degree of cooperation between bedside and remote care has been described as a success factor of telemedical interventions [10,52]. Recent literature on the impact of tele-ICU interventions suggest that effectiveness is enhanced when comanagement and clear autonomy of the remote care team are allowed [10,36].

Particularly for intensive care nurses, there is a need to establish clear rules of engagement to avoid conflicting orders between bedside and remote teams [53]. A recent ethnographic review also indicates that the perceived value of the intervention by bedside staff is a contributing factor to the success of the intervention [14]. The core research group discussed in particular the aspect of bedside physician's trust in the remote specialist. As an example, situations where an experienced physician of a nonacademic hospital in a rural area collaborates with a less experienced physician at a university hospital telemedical center can raise the issues of perceived value and trust between remote and bedside personnel. Therefore, the involvement of bedside staff during planning, system implementation, and training is recommended to enhance organizational acceptance [54,55]. As part of the implementation process, actions targeted at team cohesion (eg, team building) and use of standardized communication practices between teams can enhance the implementation of new workflows [56,57]. Implementation of health technology can lead to changes in work practice inside the care team, in particular for nursing and support staff [58]. Clear definition of the roles, responsibility, and composition of the team should therefore be addressed early on during the planning of the intervention.

Implementation of tele-ICU systems has been advanced as a solution for community hospitals facing the challenge of sustaining the cost of maintaining a local ICU with high standard of care. Economic evaluations of tele-ICU interventions are therefore an important aspect for consideration in the community settings. With tele-ICU systems, community hospitals have the potential to treat patients with a higher case mix index locally and at lower cost [51]. At the same time, high cost of tele-ICU systems has been described as a barrier to implementation [59]. Our finding indicates that studies on cost-effectiveness in this use case have not yielded uniform results. The included studies in this review have used heterogeneous approaches to estimate savings and revenue increase following tele-ICU implementation. We corroborate previous observations concerning the lack of transparency and comprehensive data on costs, which hinder comparisons and clear statements regarding cost-efficacy [59,60].

Use Case: Improving Compliance

In this use case, ICU systems were primarily configured as scheduled daily rounds from a tele-ICU center (configuration cluster *centralized scheduled interventions*; n=4, 16% of studies) and decentralized systems allowing expert remote consultations (configuration cluster *decentralized scheduled interventions*; n=5, 20% of studies). Interventions in the use case are mainly focused on advancing adherence to best practices in the ICU and increasing patient safety. They consisted in establishing critical care processes in which the remote care teams monitor relevant quality indicators (eg, prophylaxis for stress ulcer, ventilator-associated pneumonia, or deep vein thrombosis). In our analysis, there is some evidence that ICU interventions are conducive to higher adherence to best practices in the ICU and enhance patient safety, thereby corroborating earlier observations on efficacy [51]. We found that most evidence for this type of intervention has been reported in tertiary care hospitals with a closed or mixed ICU model. Additional research would be needed to understand how this type of intervention could be beneficial in other hospital contexts. The review highlighted an intervention specialized on prevention of sepsis (Deisz et al [42]), for which compliance to the care bundle was found to remain low [61,62].

We hypothesize that the efficacy of these interventions is derived from a combination of change in the care process (eg, increased use of reminders and checklists) and the use of decision support systems (eg, smart alerts). Tele-ICU systems are conducive to real-time benchmarking of performance and allow targeted actions to enhance compliance and care quality. Surveillance systems can improve resource allocation by allowing for more rapid response time and faster escalation of the most acute cases [54]. Additionally, tele-ICU systems have been shown to reduce alarm fatigue through triage and curation of automatic alerts by remote care teams [51,59]. In recent literature, the potential of population management systems allowing targeted interventions on patients with high risk factors has been highlighted [63]. Significant amount of data generated by tele-ICU systems can be leveraged for the development of advanced applications [64]. A recent systematic review on telemedicine with clinical decision support for critical care indicated the need for further research on the use and efficacy of advanced applications in units equipped with telemedical systems [65].

Use Case: Facilitating Transfer

Interventions in this use case are aimed at supporting patient transfers between hospitals (ie, referral to higher level hospital) and monitoring patients during admission in the ICU from another department (eg, emergency department). This form of intervention has been described in the literature as consultative critical care services [66]. One study in the review documented the benefit of these interventions for patients in the emergency department with suspected sepsis diagnosis [54]. Based on the studies in the review, we can corroborate previous reports that no strong evidence has been found regarding the benefit on the number of transfers for this type of interventions [67].

Limitations

Our approach has multiple limitations. First, the studies included in the review used heterogeneous research methods. Authors provided varying degree of details to describe the intervention setup and implementation context. Aspects such as staff interaction and level of autonomy have been provided only in a limited number of studies, so that our ability to draw generalizable conclusions on these aspects of tele-ICU interventions has been limited. Second, relying on the expertise of the core research group to complete the data charting was qualitative in nature and potentially subject to bias. A discussion process section was established to mitigate the interpretation bias in our approach. Third, the scope of this review was limited to the implementation of tele-ICU systems for adult patients, and critical care telemedical interventions have also been documented in pediatrics and neonatology. Some of our conclusions might therefore not be applicable to these settings.

Conclusion

Tele-ICU systems have been deployed in numerous implementation contexts, which we characterized in three main

use cases. The benefits of tele-ICU interventions have been well documented for centralized systems aimed at extending critical care capabilities in community settings and improving care compliance in tertiary hospitals. This scoping review provides teams involved in the implementation of tele-ICU systems with an overview of existing evidence on the technology. It highlights factors that are conducive to successful implementation for different critical care context. This review also mentions areas for future research on tele-ICU interventions. Furthermore, the framework for describing the implementation context used in this scoping review could be used for analyzing other types of telemedical interventions or other domains of intervention (eg, traumatology, pediatrics, neonatology).

Acknowledgments

We acknowledge support from the German Research Foundation and the Open Access Publication Funds of Charité–Universitätsmedizin Berlin. ASP is a participant in the Digital Clinician Scientist Program funded by the Charité–Universitätsmedizin Berlin and the Berlin Institute of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries for peer-reviewed studies in literature databases.
[PDF File (Adobe PDF File), 54 KB - jmir_v23i11e32264_app1.pdf]

References

1. Caples SM. Intensive care unit telemedicine care models. Crit Care Clin 2019 Jul;35(3):479-482. [doi: 10.1016/j.ccc.2019.02.004] [Medline: 31076047]
2. Becker C, Frishman WH, Scurlock C. Telemedicine and tele-ICU: the evolution and differentiation of a new medical field. Am J Med 2016 Dec;129(12):e333-e334. [doi: 10.1016/j.amjmed.2016.05.045] [Medline: 27576079]
3. Davis TM, Barden C, Dean S, Gavish A, Goliash I, Goran S, et al. American telemedicine association guidelines for TeleICU operations. Telemed J E Health 2016 Dec;22(12):971-980. [doi: 10.1089/tmj.2016.0065] [Medline: 27508454]
4. Vranas KC, Slatore CG, Kerlin MP. Telemedicine coverage of intensive care units: a narrative review. Ann Am Thorac Soc 2018 Nov;15(11):1256-1264 [FREE Full text] [doi: 10.1513/AnnalsATS.201804-225CME] [Medline: 30382786]
5. Binder WJ, Cook JL, Gramze N, Airhart S. Telemedicine in the intensive care unit: improved access to care at what cost? Crit Care Nurs Clin North Am 2018 Jun;30(2):289-296. [doi: 10.1016/j.cnc.2018.02.010] [Medline: 29724446]
6. Rogove H. How to develop a tele-ICU model? Crit Care Nurs Q 2012;35(4):357-363. [doi: 10.1097/CNQ.0b013e318266bdf5] [Medline: 22948370]
7. Hassan E. Tele-ICU and patient safety considerations. Crit Care Nurs Q 2018;41(1):47-59. [doi: 10.1097/CNQ.0000000000000185] [Medline: 29210766]
8. Welsh C, Rincon T, Berman I, Bobich T, Brindise T, Davis T. TeleICU interdisciplinary care teams. Crit Care Clin 2019 Jul;35(3):415-426. [doi: 10.1016/j.ccc.2019.02.008] [Medline: 31076042]
9. Marx G, Koch T. Telemedizin in der intensivmedizin: strukturempfehlungen der DGAI. Anästh Intensiv 2015;56:257-261.
10. Kopec IC. Impact of intensive care unit telemedicine on outcomes. Crit Care Clin 2019 Jul;35(3):439-449. [doi: 10.1016/j.ccc.2019.02.002] [Medline: 31076044]
11. Guinemer C, Boeker M, Weiss B, Fuerstenau D, Balzer F, Poncette A. Telemedicine in intensive care units: protocol for a scoping review. JMIR Res Protoc 2020 Dec 31;9(12):e19695 [FREE Full text] [doi: 10.2196/19695] [Medline: 33382040]
12. Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019.
13. Kahn JM, Le TQ, Barnato AE, Hravnak M, Kuza CC, Pike F, et al. ICU telemedicine and critical care mortality: a national effectiveness study. Med Care 2016 Mar;54(3):319-325 [FREE Full text] [doi: 10.1097/MLR.0000000000000485] [Medline: 26765148]
14. Kahn JM, Rak KJ, Kuza CC, Ashcraft LE, Barnato AE, Fleck JC, et al. Determinants of intensive care unit telemedicine effectiveness: an ethnographic study. Am J Respir Crit Care Med 2019 Apr 15;199(8):970-979 [FREE Full text] [doi: 10.1164/rccm.201802-0259OC] [Medline: 30352168]
15. Venkataraman R, Ramakrishnan N. Outcomes related to telemedicine in the intensive care unit: what we know and would like to know. Crit Care Clin 2015 Apr;31(2):225-237. [doi: 10.1016/j.ccc.2014.12.003] [Medline: 25814451]
16. Kahn JM. ICU telemedicine: from theory to practice. Crit Care Med 2014 Nov;42(11):2457-2458. [doi: 10.1097/CCM.0000000000000596] [Medline: 25319914]
17. Reynolds HN, Bander JJ. Options for tele-intensive care unit design: centralized versus decentralized and other considerations: it is not just a "another black sedan". Crit Care Clin 2015 Apr;31(2):335-350. [doi: 10.1016/j.ccc.2014.12.010] [Medline: 25814458]

18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
19. Peters MDJ, 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]
20. AHA Hospital Statistics, 2020 edition. Chicago, IL: American Hospital Association; 2020.
21. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23 [FREE Full text] [doi: 10.1197/jamia.M1749] [Medline: 16221933]
22. Breslow MJ, Rosenfeld BA, Doerfler M, Burke G, Yates G, Stone DJ, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: an alternative paradigm for intensivist staffing. Crit Care Med 2004 Jan;32(1):31-38. [doi: 10.1097/01.CCM.0000104204.61296.41] [Medline: 14707557]
23. Zawada ET, Herr P, Larson D, Fromm R, Kapaska D, Erickson D. Impact of an intensive care unit telemedicine program on a rural health care system. Postgrad Med 2009 May;121(3):160-170. [doi: 10.3810/pgm.2009.05.2016] [Medline: 19491554]
24. Thomas EJ, Lucke JF, Wueste L, Weavind L, Patel B. Association of telemedicine for remote monitoring of intensive care patients with mortality, complications, and length of stay. JAMA 2009 Dec 23;302(24):2671-2678. [doi: 10.1001/jama.2009.1902] [Medline: 20040555]
25. Franzini L, Sail KR, Thomas EJ, Wueste L. Costs and cost-effectiveness of a telemedicine intensive care unit program in 6 intensive care units in a large health care system. J Crit Care 2011 Jun;26(3):329.e1-329.e6 [FREE Full text] [doi: 10.1016/j.jcrc.2010.12.004] [Medline: 21376515]
26. Willmitch B, Golembeski S, Kim SS, Nelson LD, Gidel L. Clinical outcomes after telemedicine intensive care unit implementation. Crit Care Med 2012 Feb;40(2):450-454. [doi: 10.1097/CCM.0b013e318232d694] [Medline: 22020235]
27. Morrison JL, Cai Q, Davis N, Yan Y, Berbaum ML, Ries M, et al. Clinical and economic outcomes of the electronic intensive care unit: results from two community hospitals. Crit Care Med 2010 Jan;38(1):2-8. [doi: 10.1097/CCM.0b013e3181b78fa8] [Medline: 19730249]
28. McLeroy R, Ingersoll J, Nielsen P, Pamplin J. Implementation of tele-critical care at General Leonard Wood Army Community Hospital. Mil Med 2020 Feb 13;185(1-2):e191-e196. [doi: 10.1093/milmed/usz147] [Medline: 31247104]
29. Sadaka F, Palagiri A, Trottier S, Deibert W, Gudmestad D, Sommer SE, et al. Telemedicine intervention improves ICU outcomes. Crit Care Res Pract 2013;2013:456389. [doi: 10.1155/2013/456389] [Medline: 23365729]
30. McCambridge M, Jones K, Paxton H, Baker K, Sussman EJ, Etchason J. Association of health information technology and teleintensivist coverage with decreased mortality and ventilator use in critically ill patients. Arch Intern Med 2010 Apr 12;170(7):648-653. [doi: 10.1001/archinternmed.2010.74] [Medline: 20386011]
31. Pannu J, Sanghavi D, Sheley T, Schroeder DR, Kashyap R, Marquez A, et al. Impact of telemedicine monitoring of community ICUs on interhospital transfers. Crit Care Med 2017 Aug;45(8):1344-1351 [FREE Full text] [doi: 10.1097/CCM.0000000000002487] [Medline: 28481753]
32. Al-Omari A, Al Mutair A, Al Ammary M, Aljamaan F. A multicenter case-historical control study on short-term outcomes of tele-intensive care unit. Telemed J E Health 2020 May;26(5):645-650. [doi: 10.1089/tmj.2019.0042] [Medline: 31436501]
33. Gupta S, Dewan S, Kaushal A, Seth A, Narula J, Varma A. eICU reduces mortality in STEMI patients in resource-limited areas. Glob Heart 2014 Dec;9(4):425-427. [doi: 10.1016/j.gheart.2014.07.006] [Medline: 25592796]
34. Rosenfeld BA, Dorman T, Breslow MJ, Pronovost P, Jenckes M, Zhang N, et al. Intensive care unit telemedicine: alternate paradigm for providing continuous intensivist care. Crit Care Med 2000 Dec;28(12):3925-3931. [doi: 10.1097/00003246-200012000-00034] [Medline: 11153637]
35. Panlaqui OM, Broadfield E, Champion R, Edington JP, Kennedy S. Outcomes of telemedicine intervention in a regional intensive care unit: a before and after study. Anaesth Intensive Care 2017 Sep;45(5):605-610 [FREE Full text] [doi: 10.1177/0310057X1704500511] [Medline: 28911290]
36. Hawkins HA, Lilly CM, Kaster DA, Groves RH, Khurana H. ICU telemedicine comanagement methods and length of stay. Chest 2016 Aug;150(2):314-319. [doi: 10.1016/j.chest.2016.03.030] [Medline: 27048869]
37. Ruesch C, Mossakowski J, Forrest J, Hayes M, Jahrsdoerfer M, Comeau E, et al. Using nursing expertise and telemedicine to increase nursing collaboration and improve patient outcomes. Telemed J E Health 2012 Oct;18(8):591-595. [doi: 10.1089/tmj.2011.0274] [Medline: 22957503]
38. Lilly CM, Motzkus C, Rincon T, Cody SE, Landry K, Irwin RS, UMass Memorial Critical Care Operations Group. ICU telemedicine program financial outcomes. Chest 2017 Feb;151(2):286-297. [doi: 10.1016/j.chest.2016.11.029] [Medline: 27932050]
39. Lilly CM, Cody S, Zhao H, Landry K, Baker SP, McIlwaine J, University of Massachusetts Memorial Critical Care Operations Group. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. JAMA 2011 Jun 01;305(21):2175-2183. [doi: 10.1001/jama.2011.697] [Medline: 21576622]
40. Kohl BA, Fortino-Mullen M, Praestgaard A, Hanson CW, Dimartino J, Ochroch EA. The effect of ICU telemedicine on mortality and length of stay. J Telemed Telecare 2012 Jul;18(5):282-286. [doi: 10.1258/jtt.2012.120208] [Medline: 22802522]

41. Kalb T, Raikhelkar J, Meyer S, Ntimba F, Thuli J, Gorman MJ, et al. A multicenter population-based effectiveness study of teleintensive care unit-directed ventilator rounds demonstrating improved adherence to a protective lung strategy, decreased ventilator duration, and decreased intensive care unit mortality. J Crit Care 2014 Aug;29(4):691.e7-691.14 [FREE Full text] [doi: 10.1016/j.jcrc.2014.02.017] [Medline: 24636928]
42. Deisz R, Rademacher S, Gilger K, Jegen R, Sauerzapfe B, Fitzner C, et al. Additional telemedicine rounds as a successful performance-improvement strategy for sepsis management: observational multicenter study. J Med Internet Res 2019 Jan 15;21(1):e11161 [FREE Full text] [doi: 10.2196/11161] [Medline: 30664476]
43. Kahn JM, Gunn SR, Lorenz HL, Alvarez J, Angus DC. Impact of nurse-led remote screening and prompting for evidence-based practices in the ICU. Crit Care Med 2014 Apr;42(4):896-904 [FREE Full text] [doi: 10.1097/CCM.0000000000000052] [Medline: 24201176]
44. Vespa PM, Miller C, Hu X, Nenov V, Buxey F, Martin NA. Intensive care unit robotic telepresence facilitates rapid physician response to unstable patients and decreased cost in neurointensive care. Surg Neurol 2007 Apr;67(4):331-337. [doi: 10.1016/j.surneu.2006.12.042] [Medline: 17350395]
45. Kadar RB, Amici DR, Hesse K, Bonder A, Ries M. Impact of telemonitoring of critically ill emergency department patients awaiting ICU transfer. Crit Care Med 2019 Sep;47(9):1201-1207. [doi: 10.1097/CCM.0000000000003847] [Medline: 31162198]
46. Machado SM, Wilson EH, Elliott JO, Jordan K. Impact of a telemedicine eICU cart on sepsis management in a community hospital emergency department. J Telemed Telecare 2018 Apr;24(3):202-208. [doi: 10.1177/1357633X17691862] [Medline: 29278979]
47. Palabindala V, Abdul Salim S. Era of hospitalists. J Community Hosp Intern Med Perspect 2018;8(1):16-20 [FREE Full text] [doi: 10.1080/20009666.2017.1415102] [Medline: 29441160]
48. Sweigart JR, Aymond D, Burger A, Kelly A, Marzano N, McIlraith T, et al. Characterizing hospitalist practice and perceptions of critical care delivery. J Hosp Med 2018 Jan 01;13(1):6-12. [doi: 10.12788/jhm.2886] [Medline: 29240847]
49. Gutsche JT, Raiten JM. Staffing models for the ICU: open, closed, MD, NP, or telemedicine? Curr Anesthesiol Rep 2013 Feb 12;3(2):65-72. [doi: 10.1007/s40140-013-0010-0]
50. Hidalgo J, Pérez-Fernández J, Rodríguez-Vega G, editors. Critical Care Administration A Comprehensive Clinical Guide. Cham: Springer; 2020.
51. Fuhrman SA, Lilly CM. ICU telemedicine solutions. Clin Chest Med 2015 Sep;36(3):401-407. [doi: 10.1016/j.ccm.2015.05.004] [Medline: 26304277]
52. Wilcox ME, Adhikari NK. The effect of telemedicine in critically ill patients: systematic review and meta-analysis. Crit Care 2012 Jul 18;16(4):R127 [FREE Full text] [doi: 10.1186/cc11429] [Medline: 22809335]
53. Young LB, Chan PS, Cram P. Staff acceptance of tele-ICU coverage: a systematic review. Chest 2011 Feb;139(2):279-288 [FREE Full text] [doi: 10.1378/chest.10-1795] [Medline: 21051386]
54. Lilly CM, Mickelson JT. Evolution of the intensive care unit telemedicine value proposition. Crit Care Clin 2019 Jul;35(3):463-477. [doi: 10.1016/j.ccc.2019.02.010] [Medline: 31076046]
55. Nguyen YL, Kahn JM, Angus DC. Reorganizing adult critical care delivery: the role of regionalization, telemedicine, and community outreach. Am J Respir Crit Care Med 2010 Jun 01;181(11):1164-1169. [doi: 10.1164/rccm.200909-1441CP] [Medline: 20224067]
56. Liesching T, Lei Y. Nursing and provider roles in the tele-ICU. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019:87.
57. Ramnath VR, Malhotra A. Remote proactive physiologic monitoring in the ICU. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019:21.
58. Bergey MR, Goldsack JC, Robinson EJ. Invisible work and changing roles: health information technology implementation and reorganization of work practices for the inpatient nursing team. Soc Sci Med 2019 Aug;235:112387. [doi: 10.1016/j.socscimed.2019.112387] [Medline: 31272078]
59. Bender W, Hiddleson CA, Buchman TG. Intensive care unit telemedicine: innovations and limitations. Crit Care Clin 2019 Jul;35(3):497-509. [doi: 10.1016/j.ccc.2019.02.011] [Medline: 31076049]
60. Subramanian S, Palmer CM. Cost-benefit analysis of implementing telemedicine in the ICU. In: Telemedicine in the ICU. Cham: Springer; 2019:197.
61. Venkataraman R, Ramakrishnan N. Safety and quality metrics for ICU telemedicine: measuring success. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019:145.
62. Mohr NM, Hurst EK, MacKinney AC, Nash EC, Carr BG, Skow B. Telemedicine for early treatment of sepsis. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019.
63. Udeh C, Udeh B, Rahman N, Canfield C, Campbell J, Hata JS. Telemedicine/virtual ICU: where are we and where are we going? Methodist Debakey Cardiovasc J 2018;14(2):126-133 [FREE Full text] [doi: 10.14797/mdcj-14-2-126] [Medline: 29977469]
64. Fortis S, Goede MR. Structure and design of the tele-ICU. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019:107.

65. Mackintosh N, Terblanche M, Maharaj R, Xyrichis A, Franklin K, Keddie J, et al. Telemedicine with clinical decision support for critical care: a systematic review. Syst Rev 2016 Oct 18;5(1):176 [FREE Full text] [doi: 10.1186/s13643-016-0357-7] [Medline: 27756376]
66. Romig M, Derrett R, Latif A, Sapirstein A. Telemedicine consultation to the general ICU. In: Koenig MA, editor. Telemedicine in the ICU. Cham: Springer; 2019:307.
67. Becker C, Fusaro M, Scurlock C. Telemedicine in the ICU: clinical outcomes, economic aspects, and trainee education. Curr Opin Anaesthesiol 2019 Apr;32(2):129-135. [doi: 10.1097/ACO.0000000000000704] [Medline: 30817384]

Abbreviations

ICU: intensive care unit
PICO: Patient, Intervention, Comparison, Outcome
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by G Eysenbach; submitted 26.07.21; peer-reviewed by HS Kim, T Koritala, B Wolfe-Piggott; comments to author 26.08.21; revised version received 03.09.21; accepted 18.09.21; published 03.11.21.

Please cite as:
Guinemer C, Boeker M, Fürstenau D, Poncette AS, Weiss B, Mörgeli R, Balzer F
Telemedicine in Intensive Care Units: Scoping Review
J Med Internet Res 2021;23(11):e32264
URL: https://www.jmir.org/2021/11/e32264
doi:10.2196/32264
PMID:34730547

Review

Effectiveness of Telehealth in Rural and Remote Emergency Departments: Systematic Review

Christina Tsou[1,2*], BN, BA, MPH; Suzanne Robinson[1*], PhD; James Boyd[3*], PhD; Andrew Jamieson[2*], MBBCh, GradCertAviationMed, MOccMedHlth&Saf; Robert Blakeman[4,5*], BEc, BComPA, PGDipPsych, MCouns; Justin Yeung[6*], BMedSci, MBBS, FACEM; Josephine McDonnell[6*], BSc(Hon), MBChB, FACEM, PGradCertPEM; Stephanie Waters[2*], BSc(Physio); Kylie Bosich[6], BN, PGDipMidwifery, GAICD; Delia Hendrie[1], PhD

[1]School of Population Health, Curtin Univeristy, Bentley, Australia

[2]Innovation & Development, WA Country Health Service, Perth, Australia

[3]Digital Health, La Trobe University, Bundoora, Australia

[4]Consumer and Community Health Research Network, Nedlands, Australia

[5]Consumer and Mental Health WA, Cloverdale, Australia

[6]Command Centre, WA Country Health Service, Perth, Australia

[*]these authors contributed equally

Corresponding Author:
Christina Tsou, BN, BA, MPH
School of Population Health
Curtin Univeristy
Kent Street
Bentley, 6102
Australia
Phone: 61 08 92669266
Email: christina.tsou@curtin.edu.au

Abstract

Background: Emergency telehealth has been used to improve access of patients residing in rural and remote areas to specialist care in the hope of mitigating the significant health disparities that they experience. Patient disposition decisions in rural and remote emergency departments (EDs) can be complex and largely dependent on the expertise and experience available at local (receiving-end) hospitals. Although there has been some synthesis of evidence of the effectiveness of emergency telehealth in clinical practice in rural and remote EDs for nonacute presentations, there has been limited evaluation of the influence of contextual factors such as clinical area and acuity of presentation on these findings.

Objective: The aims of this systematic review are to examine the outcome measures used in studying the effectiveness of telehealth in rural and remote EDs and to analyze the clinical context in which these outcome measures were used and interpreted.

Methods: The search strategy used Medical Subject Headings and equivalent lists of subject descriptors to find articles covering 4 key domains: telehealth or telemedicine, EDs, effectiveness, and rural and remote. Studies were selected using the Population, Intervention, Comparator, Outcomes of Interest, and Study Design framework. This search strategy was applied to MEDLINE (Ovid), Cochrane Library, Scopus, CINAHL, ProQuest, and EconLit, as well as the Centre for Reviews and Dissemination databases (eg, National Health Service Economic Evaluation Database) for the search period from January 1, 1990, to May 23, 2020. Qualitative synthesis was performed on the outcome measures used in the included studies, in particular the clinical contexts within which they were interpreted.

Results: A total of 21 full-text articles were included for qualitative analysis. Telehealth use in rural and remote EDs demonstrated effectiveness in achieving improved or equivalent clinical effectiveness, appropriate care processes, and—depending on the context—improvement in speed of care, as well as favorable service use patterns. The definition of effectiveness varied across the clinical areas and contexts of the studies, and different measures have been used to affirm the safety and clinical effectiveness of telehealth in rural and remote EDs. The acuity of patient presentation emerged as a dominant consideration in the interpretation of interlinking time-sensitive clinical effectiveness and patient disposition measures such as transfer and discharge rates, local hospital admission, length of stay, and ED length of stay. These, together with clinical area and acuity of presentation, are the outcome determination criteria that emerged from this review.

Conclusions: Emergency telehealth studies typically use multiple outcome measures to determine the effectiveness of the services. The outcome determination criteria that emerged from this analysis are useful when defining the favorable direction for each outcome measure of interest. The findings of this review have implications for emergency telehealth service design and policies.

Trial Registration: PROSPERO CRD42019145903; https://tinyurl.com/ndmkr8ry

(J Med Internet Res 2021;23(11):e30632) doi:10.2196/30632

KEYWORDS

telehealth; telemedicine; clinical effectiveness; treatment outcome; rural population; rural health; remote

Introduction

Background

The significant health disparities for residents of rural and remote communities compared with metropolitan or urban populations raise questions of equity and access to health services. Multiple reasons have been put forward internationally to explain these health disparities [1-4]. Limited access to health care is seen as a major contributor to rural or remote and metropolitan or urban health disparities, with workforce supply central to this discourse [5-10].

Emergency department (ED) services are an essential component of the health system, often serving as the first or only point of contact for patients requiring medical care. Patients presenting to tertiary center EDs, often in an urban location, can be assured of a well-supported ED with sufficient post-ED care within the same hospital or at another hospital within a short distance [11]. Transfer decisions in rural and remote EDs do not generally have the assurance of timely and appropriate follow-up care. The challenges in rural and remote ED care are 2-fold. First, variations are evident in the capability of local hospitals, arising mainly from the lack of economy of scale to justify investments in a full range of capabilities and inpatient wards for continued treatment and monitoring after completion of ED care. Second, the time needed and distance involved to reach definitive health care compared with suburban or urban settings [12] can delay time-critical treatments such as thrombolytic treatments to resolve a dangerous clot in the blood vessels. This means that, in making decisions around patient dispositions, local hospital capabilities and distance from destination hospitals should be considered together with the patient's clinical conditions.

Emergency telehealth services provide rural and remote hospitals with timely specialist expertise to increase staff support during critical ED encounters [12], to some extent mitigating the inequities in workforce supply. The key question in the evaluation of the effectiveness of an emergency telehealth service is whether this increased specialist workforce participation in rural and remote ED presentations improves patient outcomes by delivering more timely and effective care.

Only 1 systematic review has examined the use of telehealth in rural and remote EDs, and its focus was on noncritical emergency presentations [11]. The scope of ED services included in the review ranged from telepsychiatry to assist with mental health emergency presentations and teleophthalmology for acute eye concerns requiring ophthalmologist assessment to tele-emergency, half of which involved the use of teleradiology and consultation with other subspecialists [11]. The outcomes of interest were uptake of the telehealth program, change in diagnosis or management plan, patient transfer rate, and patient dispositions (discharge, local admission, and discharge against medical advice) [11]. Of the 15 studies reviewed, 5 reported the influence of telehealth on patient diagnosis or management, with teleconsultations changing the diagnosis or management in 18%-66% of the consultations [13-17]. The review also discussed the dependence of patient dispositions on telehealth program design and observed close linkages between the rate of patient transfer, discharge, local admission, and discharge against medical advice and emergency telehealth use [11]. Most of the studies included in the review reported increases in patient transfer rates [11]. A total of 4 studies aligned telehealth with a reduction in unnecessary patient transfers [15,18-20]. Apparent in the review was the reduction of unnecessary transfers and secondary overtriage (misidentification of noncritical patients as critically ill at initial presentation), which translated into increased local hospital admissions and reduced discharge after teleconsultation [19,20].

Whether reduced unnecessary transfer and increased local admission are favorable outcomes for patients depends on the acuity and health conditions being treated as well as the infrastructure and workforce capabilities of the local hospitals to accommodate the increased demand. For the same reasons, transfer avoidance may not always result in favorable outcomes for patients presenting to rural and remote EDs. Similarly, an increase in local hospital admission may not always lead to favorable patient outcomes if specialist consultations through telehealth alone do not change the capability of local (receiving-end) hospitals to continue caring for patients presenting to the ED in critical conditions. While identifying outcome measures, the systematic review of the use of telehealth in managing emergencies in rural and remote EDs did not consider the relevance of outcome measures across the various clinical contexts such as clinical area and the acuity of presentation. In addition, its focus was on noncritical presentations only [11].

Objective

The aim of this systematic review is to examine the outcome measures used in studying the effectiveness of rural and remote emergency telehealth services and analyze the clinical context in which these outcome measures were used and interpreted. The findings from this review provide insight into evaluating the clinical impact of telehealth services in rural and remote EDs and will assist in the design of future studies on the

effectiveness and cost-effectiveness of emergency telehealth services in the rural and remote context.

Methods

Study Selection

This systematic review followed the effectiveness part of a published protocol on reviewing the literature on the effectiveness and cost-effectiveness of telehealth services in rural and remote EDs [21]. Studies were selected using the Population, Intervention, Comparator, Outcomes of Interest, and Study Design framework (see Table 1 for inclusion and exclusion criteria). Although there was a substantial body of literature on using telehealth to support prehospital emergency medical services, to restrict the scope, this review only included studies taking place in hospital EDs and excluded records reporting on the use of prehospital emergency telehealth.

Table 1. Selection criteria.

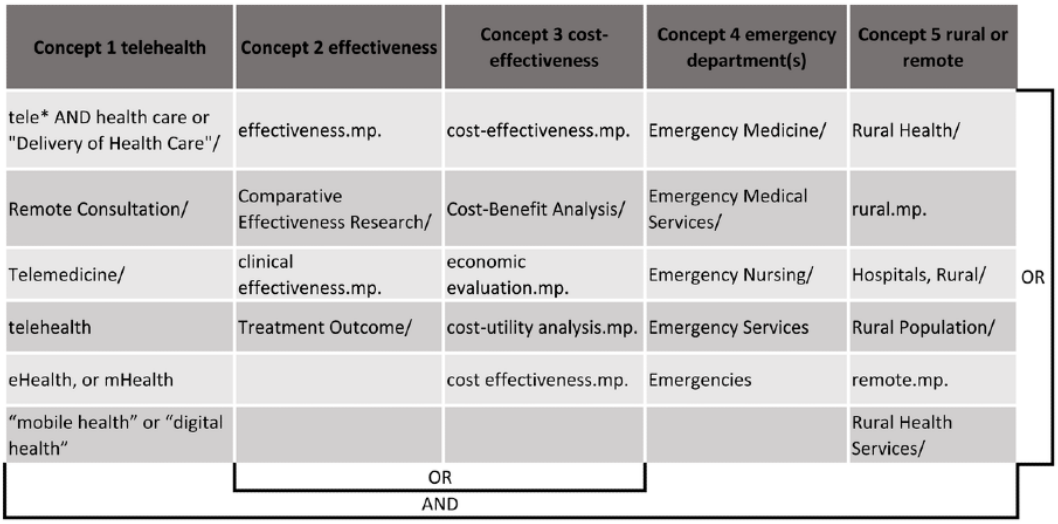
Parameter	Inclusion criteria	Exclusion criteria
Population	• Rural and remote populations treated in emergency departments	• Rural and remote populations treated outside of the emergency department
Intervention and comparator	• Emergency telehealth versus treatment as usual including the following: • Telephone versus face-to-face consults • Videoconference versus telephone consults	• Descriptive studies without comparators • Study focused on a mobile device or electronic health records
Outcomes	• Timeliness of care • Health service use • Clinical effectiveness	• Descriptive statistics without a well-defined effectiveness measure
Study design	• Randomized controlled trials • Nonrandomized controlled trials • Qualitative research	• Commentaries • Expert opinions • Government reports • Strategic documents • Single-case reports

Information Sources and Search Strategy

The search strategy used Medical Subject Headings and equivalent lists of subject descriptors to find articles covering 4 key domains: telehealth or telemedicine, EDs, effectiveness, and rural and remote (Figure 1).

This search strategy was applied to MEDLINE (Ovid), Cochrane Library, Scopus, CINAHL, ProQuest, and EconLit, as well as the Centre for Reviews and Dissemination databases (eg, the National Health Service Economic Evaluation database) for the search period from January 1, 1990, to May 23, 2020. The reference lists of the included studies were hand searched to include other peer-reviewed publications relevant to this review. Finally, a search was conducted on Google using the phrase “effectiveness of rural and remote emergency department telehealth.”

Figure 1. Search strategy.



Study Records

Data Management and Selection Process

The identified records were downloaded into EndNote (Clarivate), where duplicate records were identified and removed. A reviewer (CT) screened all the downloaded titles and excluded those that were irrelevant to the review. The abstracts of the preliminary list were classified into included, excluded, and *gray area* according to the study selection criteria [21]. The *gray area* abstract entries were independently reviewed by 1 of the 3 other reviewers (SR, DH, or JB). The articles excluded at full-text review were reviewed by another reviewer. Data extraction was organized into data collection tables, which were checked by a second reviewer (SR, DH, or JB). Any disagreements were reviewed by a third reviewer and agreed upon through discussion. The Joanna Briggs Appraisal Checklists [22] that corresponded to the study designs were used to assess the quality of the studies included for detailed review. Alignment to more than 75% of all checklist items was considered high-quality reporting, alignment to between 50% and 74% was considered moderate-quality reporting, and alignment to less than 50% was considered poor-quality reporting [23,24].

Data Extraction and Synthesis

The articles included for full-text review were categorized by clinical area, country, and operational use of telehealth interventions (whether the telehealth was used to provide direct consultation to patients at a remote end, to support local clinicians in face-to-face patient care, or for remote monitoring of changing health conditions). The outcome measures from each study were also mapped against the clinical areas to understand the use context of each type of measure. Each of the study outcomes was separately reviewed by categorizing them into clinical effectiveness or service use measures, the context in which the outcomes were used, and any validity or data quality issues noted. Data from the effectiveness studies summarized above were used to build an evidence table for each of the outcome measures identified. Salient trends were extracted from this evidence table to compile separate summaries on outcome measure use with favorable directions of change for the time-sensitive clinical effectiveness (Figure 2), service use (Figure 3), and clinical effectiveness measures (Figure 4).

Figure 2. Summary of time-sensitive clinical effectiveness.

	Duration	Length of Stay
Burns		(+*/+) Mean hospital LOS (Saffle et al)
Cardiovascular	(-*/-) FMC to wire passage through the culprit lesion (Kleinrok et al)	
	(-/-) Onset to FMC (Kleinrok et al)	
Mental health	(-*/-) Mean door to consult time (Southard et al)	(-*/-) Mean ED LOS (Southard et al)
	(-*/-) Mean order to consult time (Southard et al)	(-/-) Mean ED LOS (Vakkalanka et al)
Minor injury	(+*/0) Mean consult duration (Benger et al)	
	(+*/0) Wait time for consultation (Benger et al)	
Rural and Remote ED	(-*/-) Mean door to provider time (Mohr et al RR)	(+*/+) Mean ED LOS (Mohr et al RR)
		(0/dns) Local hospital LOS (Westbrook et al)
		(0/dns) Tertiary hospital LOS (Westbrook et al)
Stroke	(-*/-)Mean decision to rtPA time (Barlinn et al)	(0/dns) Mean hospital length of stay (Wang et al)
	(-*/-)Median door to EVT (Barlinn et al)	
	(-*/-)Median image to EVT (Barlinn et al)	
	(-*/-)Mean onset to call (Sairanen et al)	
	(-*/-)Median door to FMC (Bladin et al)	
	(-*/-)Median door to CT (Bladin et al)	
	(0/-) Mean door to call (Meyer et al)	
	(0/-) Mean door to CT (Wang et al)	
	(0/-) Mean door to decision (Meyer et al)	
	(0/-) Mean door to EKG (Meyer et al)	
	(0/-) Median door to imaging (Barlinn et al)	
	(0/-) Mean door to lab (Meyer et al; Wang et al)	
	(0/-) Mean door to FMC (Meyer et al)	
	(0/-) Mean door to tPA (Wang et al)	
	(0/-) Mean onset to door (Meyer et al)	
	(0/-) Median onset to door (Barlinn et al)	
	(0/-) Median onset to IVT (Barlinn et al)	
	(0/-) Mean onset to treatment (Sairanen et al)	
	(0/-) Median onset to arrival (Bladin et al)	
	(+*/0) Door to neurological examination time; call to neurological examination time; call to decision time; consultation duration	
Trauma	(-*/-) Transfer time (Duchesne et al)	(-*/-) Difference in ED LOS (Mohr et al T)
		(-*/-) Local hospital LOS (Duchesne et al)
Improve	No change	Worse

Figure 3. Summary of health service use measures.

	Patient disposition	Service coverage
Mental health	(+*/+) Admission rate (hospital) (Vakkalanka et al) Changes in disposition plan (disposition type) (Southard et al)	Changes in shift when consult completed (redistribution to evening and night shift); Changes in triage shift distribution (redistribution to evening and night shift) (Southard et al)
Minor injury		(+*/+) Patient asked to attend follow-up; (+*/+) Patient radiographed (Benger et al)
Paediatrics	(+*/+)Odds of transfers to a lower level care (Harvey et al) (-/-) Rate of transfer of highest acuity children to higher level care (Yang et al)	
Rural and Remote ED	(-* for critical care/-) Likelihood of local admission; (+* for moderate trauma/+) Likelihood of discharge from ED; (+*/+ for critical care; -*/- for moderate trauma) Likelihood of transfer (Westbrook et al) (-*/-)Discharge rate from ED; (+*/+) Local admission rate (Sterling et al)	(0/dns) Total ED patient volume (hospital median) (Sterling et al)
	(0/dns) Transfer rate (Sterling et al)	
	(+*/0) Discharge against medical advice (Sterling et al)	
Trauma	(0/dns) Interhospital transfer (Mohr et al T)	(0/dns) Use of diagnostic imaging (Mohr et al T)
Improve	No change	Worse

Figure 4. Summary of clinical effectiveness.

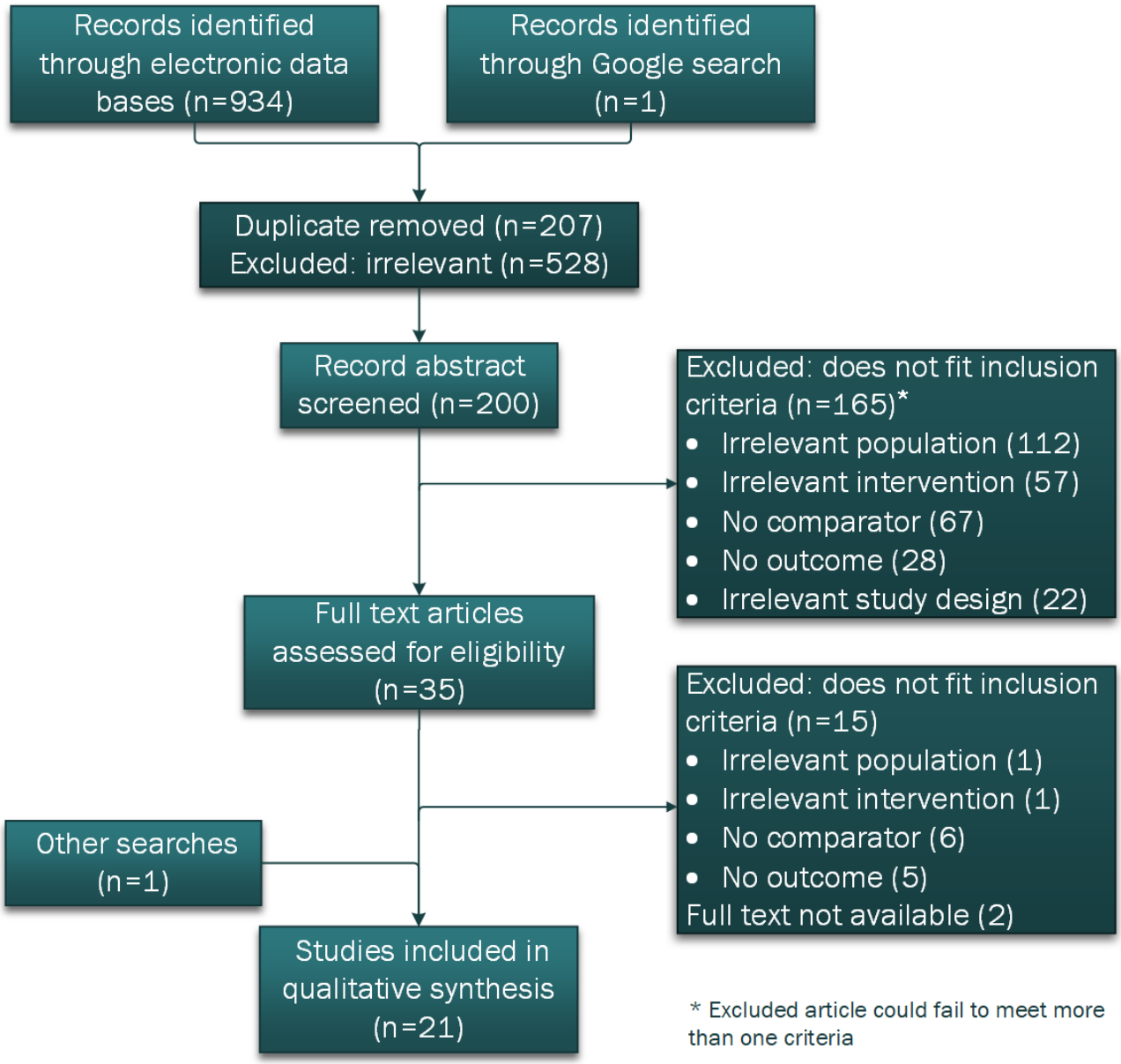
Outcome Themes	Diagnosis accuracy	Quality of consults	Treatment appropriateness	Mortality/survival	Patient outcome
Burns	(+*/+) Burn-size estimate (Saffle et al)				
Mental Health			(-*/-) Involuntary hold placement (Vakkalanka et al) (+/-) Restraint use (Vakkalanka et al)	(0/+) 6-month survival (Vakkalanka et al)	
Minor Injury			(+/0) Under treatment rate (Benger et al)		
Ophthalmology	(+*/+) Agreement on diagnosis (Bowman et al)				
Paediatrics	(+*/+) Accuracy of clinical picture (Harvey et al)	(0/0) TM improve or greatly improve quality of consult (Harvey et al)	(-*/-) Physician-related medication error (Dharmar et al)		
Rural and Remote ED				(0/0) Mortality (in-hospital) (Sterling et al)	(0/-) Change in rapid acute physiology score from time of presentation to arrival of air retrieval service (Westbrook et al)
Stroke	(+*/+) Correctness of thrombolytic treatment decision (Meyer et al)		(+/+) rtPA use rate (Meyer et al)	(0/-) Mortality (3-months) 26 28	(0/-) Intracranial bleeding (Myer et al, Barlinn et al, Sairanen et al, Yaghi et al) (0/+) Major reperfusion (Barlinn et al) (0/-) mRS score (3 months; 90-days; discharge) (Sairanen et al) (+/+) mRS >2 (3 months); (0/-) Recurrent stroke or TIA within 90 days (Yaghi et al) (+/+) mRS show good outcome (admission v discharge) (Wang et al)
			(+*/+) Odds of receiving thrombolysis among telestroke patients (Bladin et al)	(-*/-) Mortality (in-hospital) (Barlinn et al)	(+*/0) mRS >2 (3 month) patient with NIHSS > 8 (Yaghi et al)
Trauma				(0/0) Mortality (in-hospital) Duchesne et al; Mohr et al)	
Improve	No change	Worse			

Results

Study Selection

The search of the electronic databases identified 934 records for title screening. An additional record was identified from the Google search. Of the total 935 records, 207 (22.1%) duplicate titles and 528 (56.5%) irrelevant titles were removed. Of the remaining 200 abstracts screened, 165 (82.5%) were excluded because they did not fit 1 or more of the inclusion criteria. A full-text review of the remaining 35 records identified a further 15 (43%) that did not meet 1 of the inclusion criteria. An additional record was identified from hand searching of the reference lists of the included records. Detailed review and data extraction were performed on 21 articles. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram is presented in Figure 5.

Figure 5. Flow chart diagram of the study selection.



Multimedia Appendix 1 [25-36] summarizes the bibliographic information from individual records, including information related to the telehealth intervention, study details, and the authors' assessment of the reporting quality. Of the 21 studies, 14 (67%) were of high-quality reporting according to the criteria used in this review.

Study Characteristics

The studies were categorized according to clinical area and country of implementation (Table S1 of Multimedia Appendix 2 [25-36]). Of the 21 records reviewed, the most frequently assessed service (6/21, 29%) was telestroke [16,25-28,36]. Other clinical area categories included ED services in rural and remote regions [12,19,20,29], trauma [30-32], pediatrics [15,33], mental health [18,34], ophthalmology [14], minor injury [13], cardiovascular [35], and burns [17]. Of the 21 studies, 13 (62%) reported telehealth interventions implemented in the United States [12,15-18,20,27-29,31-34]; the other countries reporting on telehealth interventions were the United Kingdom [13,14], Australia [19,36], Finland [26], Germany [25], Norway [30], and Poland [35].

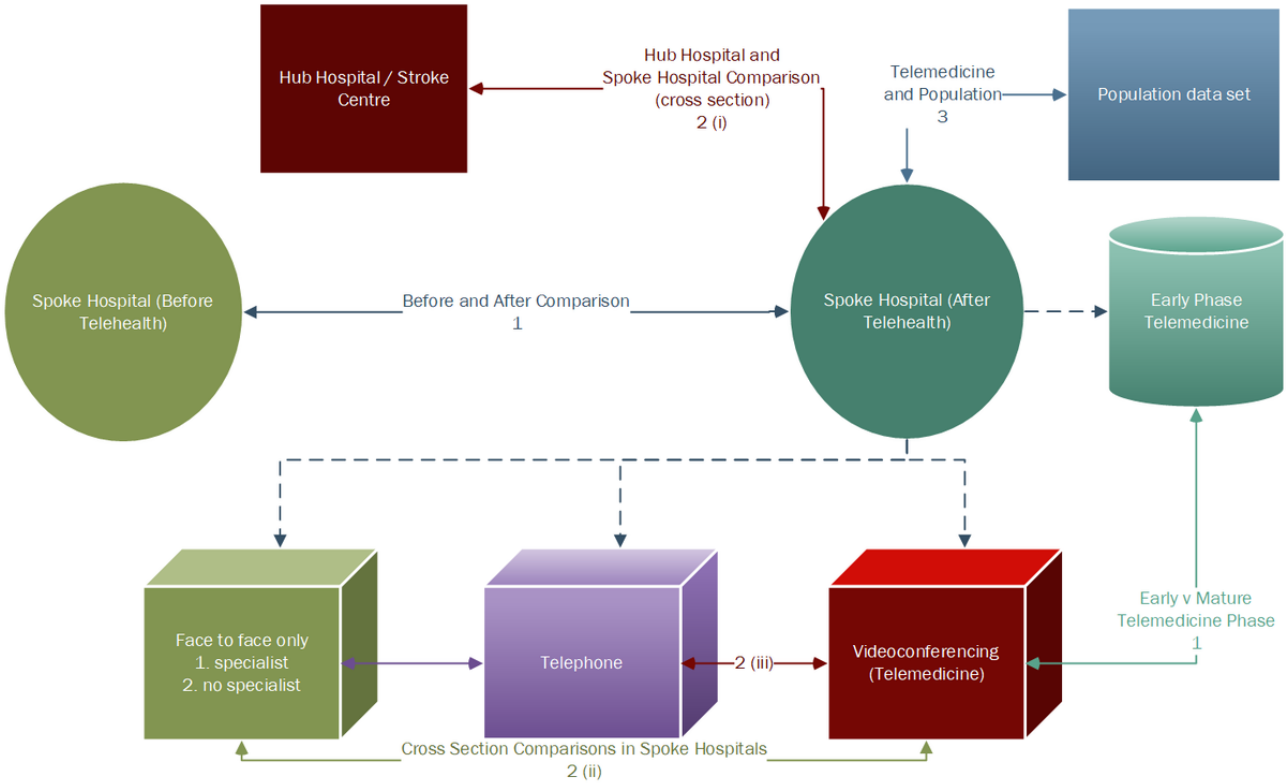
A further categorization was conducted of the clinical area by operational use of telehealth (Table S2 of Multimedia Appendix 2 [25-36]). Of the 21 studies, 12 (57%) involved a specialist supporting local clinicians [12,14,15,17,19,25-28,33,34,36], 8 (38%) involved a direct consultation by a specialist physician [13,16,18,20,29-32], and 1 (5%) concerned the provision of diagnostic services [35]. Of the 8 direct consultations, 7 (88%) were to a location staffed by a nurse practitioner, whereas 1 (12%) [18] involved a mental health specialist directly consulting the patient in the absence of a local clinician. No study documented an intervention involving remote monitoring in the rural and remote emergency setting.

Study Designs of Included Studies

The studies included in this review identified study cohorts according to whether the patients had telephone, videoconferencing, or face-to-face consultations. In terms of

the study design, 14% (3/21) of the studies were intervention studies—of these 3 studies, 2 (67%) were randomized [13,16] and 1 (33%) was nonrandomized [14])—10% (2/21) were qualitative studies [12,30], and 76% (16/21) were cohort studies [15,17-20,25-29,31-36]. There were 3 types of comparisons: (1) comparing before-and-after telehealth interventions at spoke hospitals [12,17-20,29,31] (the comparison of early-phase telemedicine implementation with the mature phase of telemedicine implementation in Kleinrok et al [35] and Bladin et al [36] is also a form of before-and-after comparison); (2) cross-sectional comparisons such as the following: (i) comparison between the videoconferencing consultation (typically referred to as telemedicine) at the spoke hospital and face-to-face consultation at the hub hospital [25,27,28], (ii) comparison within spoke hospitals between cohorts of patients who received telemedicine consultation and patients receiving only face-to-face consultations with a local clinician [13,14,32-34], and (iii) comparison between telemedicine and telephone consultations [15,16,30,33]; and (3) comparison of a telemedicine study cohort with population cohort–based data. A study on patients who had received thrombolysis compared the study group with population thrombolysis data sets [26]. Table S3 of Multimedia Appendix 2 [25-36] lists author names and year of publication against the study design categories and comparison categories. Figure 6 illustrates the different types of comparisons.

Figure 6. Comparisons used in rural and remote emergency department telehealth studies.



There was variability among the hub-and-spoke comparison studies, all of which were related to telestroke. The focus of 10% (2/21) of the studies was on patient disposition after thrombolysis: Wang et al [27] compared patients who received thrombolysis locally with patients who were transferred and then received thrombolysis at the hub, whereas Yaghi et al [28] compared the data of patients who had received thrombolysis locally, were transferred to the hub, and stayed locally. Barlinn et al [25] compared patients who had had a stroke and who had been transferred to the hub after telestroke consultation with patients who had been admitted directly to the hub.

Synthesis of Effectiveness Findings

Overview

The included quantitative studies typically used multiple measures within a study to assess effectiveness. Of the 19 quantitative studies, 13 (68%) showed improvement in 1 of the primary outcomes [14-20,27-29,31-35] and 6 (32%) in 1 of the secondary outcomes [13,15,17,18,29,34]. The effectiveness measures were categorized into time-sensitive effectiveness, service use measures, and clinical effectiveness, including patient outcome.

Time-Sensitive Effectiveness Measures

Overview

Of the 19 quantitative studies, 11 (57%) showed improvement in time-sensitive clinical effectiveness, including significant changes in ED length of stay [18,29,32], hospital length of stay [17,31], and reduction in duration of ED care processes (increased speed of care) [18,25,26,29,31,35,36]. Figure 2 summarizes the direction and findings on length of stay and care process measures by clinical area.

Length of Stay

Length of stay was a measure of clinical effectiveness used in the studies, and this has direct implication on resource use. However, a decreased length of stay was not always interpreted

as a favorable outcome across all studies. Increased hospital length of stay at the burn center was used to indicate a higher degree of diagnostic accuracy at the local hospital [17], whereas videoconferencing consultation was shown to improve the accuracy of triage [17]. Decreased local hospital length of stay was considered favorable in trauma care because of the promptness of radiologic evaluation through telehealth intervention [31], leading to faster disposition decisions. Similarly, a reduction in the ED length of stay reflected timely ED care for trauma [32] and mental health [18,34], whereas a longer ED length of stay reflected the expanded capacity of local EDs to manage patients locally [29]. Of the 19 studies, 2 (10%) showed equivalent local and tertiary hospital lengths of stay [19,27]; no study showed a worse outcome in terms of length of stay.

Care Processes

The process-of-care measures have been used to indicate service efficiency, but they are not always related to effectiveness. Improvements in care processes were shown in cardiovascular, stroke, and trauma care and were related to reaching clinical decisions and providing treatment interventions faster than usual care [16,25,26,31,35]. A significant reduction in door-to-consultation time was noted in delivering specialist mental health services [18] and when a combined rural and remote ED cohort [29] was studied. Treatment-related duration can affect clinical effectiveness in some clinical areas, for example, in the case of telestroke, time to thrombolysis (recombinant tissue plasminogen activator) or door-to–endovascular thrombectomy time. None of the studies provided direct evidence on the relationship between treatment-related durations and patient outcome.

Telehealth in rural and remote EDs made little difference to speed of care when the time gap was calculated from the point of symptom onset [16,25-27,35,36]. In the context of stroke care, earlier studies demonstrated that involving telehealth did not delay or speed up time to imaging and laboratory tests [16,25,27]. This was likely to be related to the reliance of stroke diagnosis on imaging and pre-established clinical standards in practice. However, an Australian study has shown results to the contrary where telestroke resulted in shorter door-to–computed tomography scan time and door-to-needle time for stroke thrombolysis [36]. A telestroke study showed longer time to receiving a neurology examination and reaching a clinical decision when videoconferencing was compared with telephone consultations [16]. Consult durations were longer in the clinical areas of minor injury than in face-to-face specialist consultations and among patients who had had a stroke and who received thrombolysis [13,26]; however, the duration of the consult decreased as health services became acquainted with the telehealth technology. These studies did not discuss the significance of this finding; however, a 2- to 10-minute difference in consult duration was minimal, considering the improved accuracy and appropriateness of the transfer decisions.

Health Services Use Measures

Of the 19 quantitative studies, 6 (32%) showed significant improvement in service use patterns in the clinical areas of rural and remote EDs [19,20], mental health [18,34], minor injury [13], and pediatrics [15]. In all, 5 of the service use measures were related to patient disposition, including significant change in hospital admissions [19,20,34], rate of discharge from ED [19,20], appropriateness of transfer [15,19], and changes in disposition plan [18], whereas 2 were related to significant changes in service coverage, including an increasing proportion of out-of-hour triage and consultation [18] and the proportion of patients asked to attend follow-up clinics [13]. The direction of outcome deemed favorable depended on the context of the study. Figure 3 summarizes patient disposition and service coverage measures and findings from the included studies by clinical area.

Transfer Rates

Whether to transfer a patient and to which location are important clinical decisions in rural and remote EDs. Depending on the capability and capacity of local hospitals, the acuity of presentation and the level of definitive care sought by the transfer are directly related to the decisions to transfer or stay locally. Transfer rates were interpreted together with admission and discharge rates and had the function of examining the appropriateness of service use. An acuity subgroup analysis in Westbrook et al [19] demonstrated significant variation in transfer rates between critical care and moderate trauma. That is, telehealth increases the transfer rate of patients classified as high acuity, which is mirrored in the reduced likelihood of local admission in this cohort. Similarly, the significant reduction in transfer likelihood of patients with moderate trauma is reflected in the increased likelihood of discharge from local EDs [37]. Pediatric emergency telehealth studies reported increased odds of transfer to lower-level care in pediatric triage [15] and reduced transfer in pediatric patients with the highest-acuity ED presentations compared with telephone consultations [38]. A study interpreted the observations on rates of transfer together with the clinicians' subjective perception of increased accuracy of the clinical picture before arrival at the tertiary hospital [15], whereas another study regarded changes in transfer rates as an indication of health services use appropriateness [19].

Local Admissions and Discharge From ED

Patients who are not transferred are either admitted locally or discharged home. This is another decision made by ED clinicians in consultation with emergency medicine specialists through telehealth. The reduction in the rates of discharge from local EDs and increased local admission was observed in Sterling et al [20] after the implementation of emergency medicine specialist consultation with a local nurse practitioner. This study did not stratify by acuity of presentation, and the higher rate of local admission explained the decreased ED discharge rate [20]. The change in discharge rate accompanying increased local admission is contrary to that reported in Westbrook et al [19], which involved decision-making by local ED physicians in consultation with emergency medicine specialists.

A reduced likelihood of local admission for patients classified as critical care and increased likelihood of discharge from the ED for patients with moderate trauma in Westbrook et al [19] indicated the effectiveness of the intervention in identifying patients who did not require further care, whereas in Sterling

et al [20], increased local admission rate and decreased rate of discharge from the ED were used to reflect the benefit of telemedicine in augmenting local capacity to care for patients locally, and this translated into increased financial viability of local hospitals. The patients in Westbrook et al [19] were classified as triage category 1 or were those with major or moderate trauma, skewing toward higher acuity compared with the Sterling et al [20] patient cohort. The variation in patient acuity reflected the base local hospital capability that the emergency medicine specialists were supporting. This demonstrated the differing use of these measures across patient acuity levels.

In addition, the different patterns of changes in the rates of transfer and discharge from the ED between the Westbrook et al [19] and Sterling et al [20] studies are not likely to be attributable to telehealth consultations. It is more likely a result of the difference in disposition practices for patients who stayed locally. Spoke hospitals in Westbrook et al [19] may have cared for and discharged patients from EDs, whereas in Sterling et al [20], the patients may have been transferred to the inpatient department for the same care, which is why they were not considered discharged from the ED.

For mental health ED presentations, an increased combined rate of hospital admissions [34] and an increased range of dispositions [18] were considered a favorable outcome from telehealth interventions. Sterling et al [20] showed an increased rate of discharge against medical advice after implementation of a telehealth service. In other words, when the treating physician was not physically present, patients were more likely to act against medical advice.

Service Coverage

Telemedicine was effective in redistributing or increasing service coverage in the ED. Southard et al [18] demonstrated a redistribution of mental health triage and consultation completed in the evening and night shifts, and Benger et al [13] showed an increase in the proportion of patients with minor injuries asked to attend follow-up appointments after telemedicine consultation. Other service coverage measures included the proportion of patients radiographed for minor injury [13], median total ED patient volume [20], and the odds of using diagnostic imaging in patients with trauma [32], but the studies did not show significant change in these service coverage measures.

Clinical Effectiveness Measures

Clinical effectiveness is an indication of safety and quality and can be used as a surrogate measure for patient outcome. Of the 19 quantitative studies, 9 (47%) showed improvement in clinical effectiveness, including diagnosis accuracy [14,15,17], treatment appropriateness [16,33,34], and improved patient outcome [25,27,28], whereas 15 (71%) showed that telemedicine can achieve effectiveness similar to that achieved by the comparator interventions. The effectiveness measures included in-hospital mortality [20,31,32,36], 3-month [26,28] and 6-month [34] mortality, treatment complications [16,25,26,28,36], patient outcomes [16,19,25-28], treatment rates [13,16,34], and consult quality [15].

The measures for diagnosis accuracy and patient outcome were clinical area dependent. A generalized study on telehealth in rural and remote EDs used change in the rapid acute physiology score from time of presentation to arrival of air retrieval service as a measure of stability among transferred patients [19]. Apart from the rapid acute physiology score, all other patient outcome measures were stroke related (ie, functional scores, major reperfusion after thrombolysis, intracranial hemorrhage, and recurrent stroke rates) [16,25-28]. Figure 4 summarizes the categories of the clinical effectiveness measure by clinical area, with suggested outcomes hierarchy among these measures. As reported above, treatment-related, time-sensitive clinical effectiveness measures can also be surrogate measures for patient outcomes.

Risk of Bias in Individual Studies

The most significant risk of bias within the studies was related to study design. The risk of bias is high in before-and-after comparison studies because systems change over time, which cannot be controlled [20]. The studies did not report on the details of system change; therefore, this risk of bias could not be addressed in this review. Study designs involving cross-sectional comparisons could not be randomized and experienced selection bias pertaining to severity, with higher severity among the telemedicine groups, that is, tendency to consult emergency medicine specialists when managing patients classified as more severe.

The tendency to recruit greater number of telemedicine cases from larger spoke hospitals diluted the effect of remoteness on the effectiveness of telehealth when the results were interpreted as a whole. This is a significant selection bias in telehealth studies and highlights the gap in the published literature on the impact of telehealth in remote and very remote regions.

Other risks of bias in individual studies related to the sample sizes. Of the 21 studies, 7 (33%) had sample sizes of fewer than 100 patients [14,16-19,26,27]. The studies also reported selection bias in relation to local (receiving-end) clinicians' preselected cases for telehealth consultations when hospital transfers were considered. This process selected patients whose transfers were more imminent after telemedicine specialist consultations.

A further weakness in many of the included studies was that of attribution. The studies assumed the clinical or service use effectiveness to be associated with the telehealth interventions without considering potential confounders.

Discussion

Principal Findings

Context of Intervention

Telehealth interventions were considered effective when their implementation resulted in improvement or equivalent clinical or service use outcomes. The indicators used to measure favorable patient outcome were unidirectional; that is, better outcome pertains to change of the indicator in 1 direction. However, time-sensitive effectiveness and service use measures

were interpreted differently, depending on the context of the intervention.

Patient Dispositions

Patient disposition measures were dependent on the severity of the presenting illness and the level of definitive care compared with the hospital of origin and clinical area. Higher rate of transfer and shorter local ED or hospital length of stay were considered favorable in higher acuity ED presentations such as triage 1 or major trauma, and the reverse was true in less life-threatening conditions such as mental health and moderate trauma. This concurs with the findings from the observational study conducted on the North Dakota critical-access hospital ED cohorts where the interhospital transfer rate was not associated with telehealth use after adjusting for severity of illness [39]. The pediatric emergency telehealth studies considered an additional factor of the relative level of definitive care. Transfers to lower-level care increased with telehealth consultations [15], whereas transfers to higher-level care decreased in pediatric cases with the highest acuity [38].

Timeliness of Care Measures

Telestroke was effective in maintaining the same level of timeliness once a patient arrived at the ED, whereas in other clinical areas, telehealth resulted in faster transitions from consultation to diagnosis and treatment. Timeliness of the care measures reflected factors modifiable by telehealth. For example, the significant reduction of door-to-consultation time may be explained by the shortage of workforce and specialist skills in the rural and remote EDs, which were modifiable using telehealth consultations. The observation in telestroke—no significant change in timeliness of care—was explained by the dependence on access to imaging and other diagnostic tests, which were local processes that were not modifiable by telehealth consultations. The competent use of technology was also a factor influencing time-sensitive clinical effectiveness.

Changing Service Use Patterns

Favorable service use patterns involved a redistribution of resources, such as changes in disposition plan, changes in the pattern of interhospital transfers [19,20,32,37,38], or redistribution of ED triage and medical consultation to evening and night shifts instead of a concentration of medical consultations during day shifts [18]. These changes indicated the effectiveness of telehealth in facilitating more appropriate use of health resources, which did not always translate into an absolute reduction of service use or cost savings for the health systems. Furthermore, transfer rates stratified by acuity of presentation and their impact on local hospital admission and ED discharge rates are also a meaningful service use measure for emergency telehealth service evaluation. This also indicates the effectiveness of specialist consultation in changing resource use patterns at a local level.

The synthesis from this review indicated that favorable health service use patterns can be expected across different clinical areas where the impact of a specialist workforce shortage is modifiable by telehealth implementation in these settings. Dorsey et al [40] observed 3 interlinking trends shaping telehealth, including the transformation from increasing access to eventually reducing cost rationalizing the potential for telehealth to reduce time spent accessing specialist services and increase the intensity of services to the 20% of the people accounting for 80% of the health expenditure. Changes in cost were not considered in this review, and the synthesis in this systematic review did not find conclusive evidence to support this trend in the context of acute presentation to rural and remote EDs. However, the increasing local hospital admission rate due to the addition of clinical expertise through emergency telehealth has been discussed in a study as a means of increasing local hospital revenue [20]. The changes in patient disposition, such as reduction in admissions and transfers [19], also align with the changes in the cost profile of the overall service delivery.

Patient Outcomes

Studies reporting clinical effectiveness demonstrated improved clinical effectiveness in stroke, pediatrics, burns, mental health, and ophthalmology, albeit by using different clinical or patient outcome measures. We have identified a hierarchy of outcomes in the studies: treatment-related timeliness, diagnosis accuracy, quality of consults, and treatment appropriateness are all categories of outcomes with potential impact on patients' functional outcome or survival.

This review was conducted as part of a larger study on the cost-effectiveness of telehealth in rural and remote EDs [21]. Patient outcomes are the foundation of cost-effectiveness analyses in rural and remote EDs. The ideal outcome measure to accommodate the wide range of ED presentations is quality-adjusted life years (QALYs). When direct data collection on QALYs is not a pragmatic option, it is often calculated from a patient outcome measure. In telestroke studies, for example, QALYs are derived from the modified Rankin Scale scores. A further extension to this review on effectiveness measures in rural and remote emergency telehealth services is the question of using mortality and patient outcome measures to derive QALYs for economic evaluation. This calls for future research into the relationship between diagnosis accuracy and mortality and functional outcomes and the derivation of QALYs from the appropriate patient outcome measures for rural and remote EDs and the receiving-end (local) hospital context of the emergency telehealth service.

Limitations

Timeliness of the clinical decision and, where appropriate, clinical intervention is critical to the effectiveness of acute ED care. In this review, 2 components that contribute to timeliness were not considered: the impact of prehospital emergency medical services and the distance factor. The decision to exclude studies on prehospital emergency medical services was made to restrict the scope of this review to the effectiveness of emergency telehealth services delivered in hospital settings.

The second limitation is related to the distance between the EDs and the destination of transfer. This, combined with the appropriateness of interhospital transfer decisions, contributes to treatment delays in transit and may affect further patient care (clinical) decisions. We were unable to determine the extent to which telehealth was effective in bridging this gap across clinical areas from the evidence reviewed. Although most studies

reported on the distance between spoke and hub hospitals, it was not possible to synthesize the impact of distance on the effectiveness findings using the available information. The small number of spokes included in the studies also made it difficult to make meaningful comparisons in relation to the distance in the studies.

The heterogeneity of the studies in this field rendered the observation of relationships among different levels in the outcomes hierarchy an impossible task. However, based on the findings in this systematic review, future studies or evaluation efforts are well placed to consider an outcomes hierarchy, with surrogate outcomes leading to changes in patients' functional outcome and mortality in the same study.

Comparison With Prior Work

This review is the first in rural and remote emergency telehealth to demonstrate the importance of understanding the context in which effectiveness measures are applied in evaluating telehealth. In designing telehealth services in rural and remote EDs, the purpose of telehealth by clinical area and acuity of presentation should be ascertained before setting targets for the telehealth services or program. The context around the benefit of telehealth in supporting more informed clinical decisions and accurate diagnosis, more favorable health service use patterns, and longer-term patient outcomes also has policy implications.

Policy makers should be cognizant of the complexities around, and the limitations of, emergency telehealth in rural and remote settings so as to set reasonable expectations regarding the expected outcomes from this modality of service delivery. The ascertainment of service goals should commence by examining the purpose of telehealth by clinical area, acuity of presentation, receiving-end hospital capability, and the level of definitive care to set appropriate performance targets.

Conclusions

Ascertaining outcome measures to accurately reflect the contribution of telehealth in rural and remote EDs is a complex task. Emergency telehealth studies typically use multiple outcome measures to determine the effectiveness of the services. The analysis in this systematic review has revealed 3 criteria in outcome determination in this context: clinical area, acuity of presentation, and the level of definitive care relative to the hospital of origin. These criteria are useful when defining the favorable direction for each outcome measure of interest.

The findings from this review inform the motivation and expectation of emergency rural and remote telehealth services in the design phase. The evidence from this review indicates that emergency telehealth service adoption has resulted in better service use patterns by improving the diagnosis and making first-line management modifiable by bringing in specialist expertise in emergency medicine. However, the factors that influence clinical decisions but are not modifiable by emergency telehealth, such as receiving-end hospital capability, have not been directly studied in the rural and remote ED context.

Acknowledgments

This research was initiated by the WA Country Health Service, Curtin University, and received financial support from both institutions as well as the Digital Health Collaborative Research Centre in the form of PhD scholarships. WA Country Health Service also contributed in-kind support in the form of clinician time, academic supervisors, and administrative and research infrastructure support.

Authors' Contributions

AJ, JY, and JM ensured relevance of the review to patient care in rural and remote emergency department settings. KB and SW ensured the relevance of this review to emergency telehealth service design, implementation, and evaluation. RB provided input from a health consumer's perspective. CT and DH conceptualized and drafted the manuscript with support from SR and JB in terms of the processing of articles downloaded for the systematic review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary information of the included studies.
[PDF File (Adobe PDF File), 166 KB - jmir_v23i11e30632_app1.pdf]

Multimedia Appendix 2
Categorization of the included studies.
[PDF File (Adobe PDF File), 132 KB - jmir_v23i11e30632_app2.pdf]

References

1. Phillips A. Health status differentials across rural and remote Australia. Aust J Rural Health 2009 Feb;17(1):2-9. [doi: 10.1111/j.1440-1584.2008.01029.x] [Medline: 19161493]

2. Liu Y, Hsiao WC, Eggleston K. Equity in health and health care: the Chinese experience. Soc Sci Med 1999 Nov;49(10):1349-1356. [doi: 10.1016/s0277-9536(99)00207-5] [Medline: 10509825]
3. Hanlon N, Kearns R. Health and rural places. In: Routledge International Handbook of Rural Studies. Oxfordshire: Routledge; 2016.
4. 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]
5. Blaauw D, Erasmus E, Pagaiya N, Tangcharoensathein V, Mullei K, Mudhune S, et al. Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment. Bull World Health Organ 2010 May;88(5):350-356 [FREE Full text] [doi: 10.2471/BLT.09.072918] [Medline: 20461141]
6. Kahabuka C, Kvåle G, Moland KM, Hinderaker SG. Why caretakers bypass Primary Health Care facilities for child care - a case from rural Tanzania. BMC Health Serv Res 2011 Nov 17;11:315 [FREE Full text] [doi: 10.1186/1472-6963-11-315] [Medline: 22094076]
7. Lori JR, Livingston L, Eagle M, Rominski S, Nakua EK, Agyei-Baffour P. Rural origin and exposure drives Ghanaian midwives reported future practice. Afr J Reprod Health 2014 Sep;18(3):95-100 [FREE Full text] [Medline: 25438514]
8. O'Donnell O. Access to health care in developing countries: breaking down demand side barriers. Cad Saude Publica 2007 Dec;23(12):2820-2834 [FREE Full text] [doi: 10.1590/s0102-311x2007001200003] [Medline: 18157324]
9. Reid SJ. Rural health and transformation in South Africa. S Afr Med J 2006 Aug;96(8):676-677. [Medline: 17019486]
10. Dolea C, World Health Organization. Increasing Access to Health Workers in Remote and Rural Areas Through Improved Retention: Global Policy Recommendations. Geneva: World Health Organization; 2010.
11. du Toit M, Malau-Aduli B, Vangaveti V, Sabesan S, Ray RA. Use of telehealth in the management of non-critical emergencies in rural or remote emergency departments: a systematic review. J Telemed Telecare 2019 Jan;25(1):3-16. [doi: 10.1177/1357633X17734239] [Medline: 28980853]
12. Mueller KJ, Potter AJ, MacKinney AC, Ward MM. Lessons from tele-emergency: improving care quality and health outcomes by expanding support for rural care systems. Health Aff (Millwood) 2014 Feb;33(2):228-234. [doi: 10.1377/hlthaff.2013.1016] [Medline: 24493765]
13. Benger JR, Noble SM, Coast J, Kendall JM. The safety and effectiveness of minor injuries telemedicine. Emerg Med J 2004 Jul;21(4):438-445 [FREE Full text] [Medline: 15208226]
14. Bowman RJ, Kennedy C, Kirwan JF, Sze P, Murdoch IE. Reliability of telemedicine for diagnosing and managing eye problems in accident and emergency departments. Eye (Lond) 2003 Aug;17(6):743-746. [doi: 10.1038/sj.eye.6700489] [Medline: 12928688]
15. Harvey JB, Yeager BE, Cramer C, Wheeler D, McSwain SD. The impact of telemedicine on pediatric critical care triage. Pediatr Crit Care Med 2017 Nov;18(11):e555-e560. [doi: 10.1097/PCC.0000000000001330] [Medline: 28922271]
16. Meyer BC, Raman R, Hemmen T, Obler R, Zivin JA, Rao R, et al. Efficacy of site-independent telemedicine in the STRokE DOC trial: a randomised, blinded, prospective study. Lancet Neurol 2008 Sep;7(9):787-795 [FREE Full text] [doi: 10.1016/S1474-4422(08)70171-6] [Medline: 18676180]
17. Saffle J, Edelman L, Theurer L, Morris S, Cochran A. Telemedicine evaluation of acute burns is accurate and cost-effective. J Trauma 2009 Aug;67(2):358-365. [doi: 10.1097/TA.0b013e3181ae9b02] [Medline: 19667890]
18. Southard EP, Neufeld JD, Laws S. Telemental health evaluations enhance access and efficiency in a critical access hospital emergency department. Telemed J E Health 2014 Jul;20(7):664-668. [doi: 10.1089/tmj.2013.0257] [Medline: 24811858]
19. Westbrook JI, Coiera EW, Brear M, Stapleton S, Rob MI, Murphy M, et al. Impact of an ultrabroadband emergency department telemedicine system on the care of acutely ill patients and clinicians' work. Med J Aust 2008 Jun 16;188(12):704-708. [doi: 10.5694/j.1326-5377.2008.tb01850.x] [Medline: 18558892]
20. Sterling SA, Seals SR, Jones AE, King MH, Galli RL, Isom KC, et al. The impact of the TelEmergency program on rural emergency care: an implementation study. J Telemed Telecare 2017 Jul;23(6):588-594. [doi: 10.1177/1357633X16657499] [Medline: 27470505]
21. Tsou C, Robinson S, Boyd J, Jamieson A, Blakeman R, Bosich K, et al. Effectiveness and cost-effectiveness of telehealth in rural and remote emergency departments: a systematic review protocol. Syst Rev 2020 Apr 17;9(1):82 [FREE Full text] [doi: 10.1186/s13643-020-01349-y] [Medline: 32303253]
22. Critical appraisal tools. Joanna Briggs Institute. URL: https://joannabriggs.org/critical_appraisal_tools [accessed 2021-11-02]
23. Yuniar P, Robinson S, Moorin R, Norman R. Economic Evaluation of Breast Cancer Early Detection Strategies in Asia: A Systematic Review. Value in Health Regional Issues 2020 May;21:252-263. [doi: 10.1016/j.vhri.2020.01.003]
24. Mangham-Jefferies L, Pitt C, Cousens S, Mills A, Schellenberg J. Cost-effectiveness of strategies to improve the utilization and provision of maternal and newborn health care in low-income and lower-middle-income countries: a systematic review. BMC Pregnancy Childbirth 2014 Jul 22;14:243 [FREE Full text] [doi: 10.1186/1471-2393-14-243] [Medline: 25052536]
25. Barlinn J, Gerber J, Barlinn K, Pallesen LP, Siepmann T, Zerna C, et al. Acute endovascular treatment delivery to ischemic stroke patients transferred within a telestroke network: a retrospective observational study. Int J Stroke 2017 Jul;12(5):502-509. [doi: 10.1177/1747493016681018] [Medline: 27899742]

26. Sairanen T, Soinila S, Nikkanen M, Rantanen K, Mustanoja S, Färkkilä M, Finnish Telestroke Task Force. Two years of Finnish Telestroke: thrombolysis at spokes equal to that at the hub. Neurology 2011 Mar 29;76(13):1145-1152. [doi: 10.1212/WNL.0b013e318212a8d4] [Medline: 21368283]
27. Wang DZ, Rose JA, Honings DS, Garwacki DJ, Milbrandt JC. Treating acute stroke patients with intravenous tPA. The OSF stroke network experience. Stroke 2000 Jan;31(1):77-81. [doi: 10.1161/01.str.31.1.77] [Medline: 10625719]
28. Yaghi S, Harik SI, Hinduja A, Bianchi N, Johnson DM, Keyrouz SG. Post t-PA transfer to hub improves outcome of moderate to severe ischemic stroke patients. J Telemed Telecare 2015 Oct;21(7):396-399. [doi: 10.1177/1357633X15577531] [Medline: 25962653]
29. Mohr NM, Young T, Harland KK, Skow B, Wittrock A, Bell A, et al. Emergency department telemedicine shortens rural time-to-provider and emergency department transfer times. Telemed J E Health 2018 Aug;24(8):582-593. [doi: 10.1089/tmj.2017.0262] [Medline: 29293413]
30. Bolle SR, Larsen F, Hagen O, Gilbert M. Video conferencing versus telephone calls for team work across hospitals: a qualitative study on simulated emergencies. BMC Emerg Med 2009 Nov 30;9:22 [FREE Full text] [doi: 10.1186/1471-227X-9-22] [Medline: 19943978]
31. Duchesne JC, Kyle A, Simmons J, Islam S, Schmieg RE, Olivier J, et al. Impact of telemedicine upon rural trauma care. J Trauma 2008 Jan;64(1):92-7; discussion 97. [doi: 10.1097/TA.0b013e31815dd4c4] [Medline: 18188104]
32. Mohr NM, Vakkalanka JP, Harland KK, Bell A, Skow B, Shane DM, et al. Telemedicine use decreases rural emergency department length of stay for transferred north Dakota trauma patients. Telemed J E Health 2018 Mar;24(3):194-202. [doi: 10.1089/tmj.2017.0083] [Medline: 28731843]
33. Dharmar M, Kuppermann N, Romano PS, Yang NH, Nesbitt TS, Phan J, et al. Telemedicine consultations and medication errors in rural emergency departments. Pediatrics 2013 Dec;132(6):1090-1097. [doi: 10.1542/peds.2013-1374] [Medline: 24276844]
34. Vakkalanka JP, Harland KK, Wittrock A, Schmidt M, Mack L, Nipe M, et al. Telemedicine is associated with rapid transfer and fewer involuntary holds among patients presenting with suicidal ideation in rural hospitals: a propensity matched cohort study. J Epidemiol Community Health 2019 Nov;73(11):1033-1039 [FREE Full text] [doi: 10.1136/jech-2019-212623] [Medline: 31492762]
35. Kleinrok A, Płaczkiewicz DT, Puźniak M, Dąbrowski P, Adamczyk T. Electrocardiogram teletransmission and teleconsultation: essential elements of the organisation of medical care for patients with ST segment elevation myocardial infarction: a single centre experience. Kardiol Pol 2014;72(4):345-354 [FREE Full text] [doi: 10.5603/KP.a2013.0352] [Medline: 24408066]
36. Bladin C, Kim J, Bagot K, Vu M, Moloczij N, Denisenko S, et al. Improving acute stroke care in regional hospitals: clinical evaluation of the Victorian Stroke Telemedicine program. Med J Aust 2020 May;212(8):371-377. [doi: 10.5694/mja2.50570] [Medline: 32255520]
37. Natafgi N, Shane DM, Ullrich F, MacKinney AC, Bell A, Ward MM. Using tele-emergency to avoid patient transfers in rural emergency departments: an assessment of costs and benefits. J Telemed Telecare 2018 Apr;24(3):193-201. [doi: 10.1177/1357633X17696585] [Medline: 29278984]
38. Yang NH, Dharmar M, Yoo BK, Leigh JP, Kuppermann N, Romano PS, et al. Economic evaluation of pediatric telemedicine consultations to rural emergency departments. Med Decis Making 2015 Aug;35(6):773-783. [doi: 10.1177/0272989X15584916] [Medline: 25952744]
39. Mohr NM, Harland KK, Chrischilles EA, Bell A, Shane DM, Ward MM. Emergency department telemedicine is used for more severely injured rural trauma patients, but does not decrease transfer: a cohort study. Acad Emerg Med 2017 Feb;24(2):177-185 [FREE Full text] [doi: 10.1111/acem.13120] [Medline: 28187248]
40. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]

Abbreviations

ED: emergency department
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY: quality-adjusted life year

Edited by R Kukafka; submitted 22.05.21; peer-reviewed by T McGough, R Ciorap; comments to author 30.07.21; revised version received 24.08.21; accepted 12.09.21; published 26.11.21.

Please cite as:
Tsou C, Robinson S, Boyd J, Jamieson A, Blakeman R, Yeung J, McDonnell J, Waters S, Bosich K, Hendrie D
Effectiveness of Telehealth in Rural and Remote Emergency Departments: Systematic Review
J Med Internet Res 2021;23(11):e30632
URL: https://www.jmir.org/2021/11/e30632
doi:10.2196/30632
PMID:34842537

Original Paper

Data Quality of Longitudinally Collected Patient-Reported Outcomes After Thoracic Surgery: Comparison of Paper- and Web-Based Assessments

Hongfan Yu[1], BSc; Qingsong Yu[1], BSc; Yuxian Nie[2], ME; Wei Xu[1], MSc; Yang Pu[1], BSc; Wei Dai[3], MD; Xing Wei[3], MD; Qiuling Shi[1,2,3], MD, PhD

[1]School of Public Health and Management, Chongqing Medical University, Chonqqing, China

[2]State Key Laboratory of Ultrasound in Medicine and Engineering, College of Biomedical Engineering, Chongqing Medical University, Chongqing, China

[3]Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

Corresponding Author:
Qiuling Shi, MD, PhD
School of Public Health and Management
Chongqing Medical University
No 1, Medical College Road, Yuzhong District
Chonqqing, 400016
China
Phone: 86 18290585397
Fax: 86 28 85420116
Email: qshi@cqmu.edu.cn

Abstract

Background: High-frequency patient-reported outcome (PRO) assessments are used to measure patients' symptoms after surgery for surgical research; however, the quality of those longitudinal PRO data has seldom been discussed.

Objective: The aim of this study was to determine data quality-influencing factors and to profile error trajectories of data longitudinally collected via paper-and-pencil (P&P) or web-based assessment (electronic PRO [ePRO]) after thoracic surgery.

Methods: We extracted longitudinal PRO data with 678 patients scheduled for lung surgery from an observational study (n=512) and a randomized clinical trial (n=166) on the evaluation of different perioperative care strategies. PROs were assessed by the MD Anderson Symptom Inventory Lung Cancer Module and single-item Quality of Life Scale before surgery and then daily after surgery until discharge or up to 14 days of hospitalization. Patient compliance and data error were identified and compared between P&P and ePRO. Generalized estimating equations model and 2-piecewise model were used to describe trajectories of error incidence over time and to identify the risk factors.

Results: Among 678 patients, 629 with at least 2 PRO assessments, 440 completed 3347 P&P assessments and 189 completed 1291 ePRO assessments. In total, 49.4% of patients had at least one error, including (1) missing items (64.69%, 1070/1654), (2) modifications without signatures (27.99%, 463/1654), (3) selection of multiple options (3.02%, 50/1654), (4) missing patient signatures (2.54%, 42/1654), (5) missing researcher signatures (1.45%, 24/1654), and (6) missing completion dates (0.30%, 5/1654). Patients who completed ePRO had fewer errors than those who completed P&P assessments (ePRO: 30.2% [57/189] vs. P&P: 57.7% [254/440]; P<.001). Compared with ePRO patients, those using P&P were older, less educated, and sicker. Common risk factors of having errors were a lower education level (P&P: odds ratio [OR] 1.39, 95% CI 1.20-1.62; P<.001; ePRO: OR 1.82, 95% CI 1.22-2.72; P=.003), treated in a provincial hospital (P&P: OR 3.34, 95% CI 2.10-5.33; P<.001; ePRO: OR 4.73, 95% CI 2.18-10.25; P<.001), and with severe disease (P&P: OR 1.63, 95% CI 1.33-1.99; P<.001; ePRO: OR 2.70, 95% CI 1.53-4.75; P<.001). Errors peaked on postoperative day (POD) 1 for P&P, and on POD 2 for ePRO.

Conclusions: It is possible to improve data quality of longitudinally collected PRO through ePRO, compared with P&P. However, ePRO-related sampling bias needs to be considered when designing clinical research using longitudinal PROs as major outcomes.

(J Med Internet Res 2021;23(11):e28915) doi:10.2196/28915

KEYWORDS

patient-reported outcome (PRO); data quality; MDASI-LC; postoperative care; symptoms

Introduction

Scientific Background

Patient-reported outcomes (PROs) are commonly assessed as primary or secondary outcomes in clinical trials or observational studies to evaluate the effect of medical interventions from the viewpoint of patients without interpretation by professionals [1-3]. PROs can help clinicians monitor adverse events [4], relieve symptom burdens [5], guide clinical care [4,6], and improve patient outcomes [6], such as quality of life (QOL) and survival. However, PROs involve multiple self-evaluations over time, and symptoms change frequently over the course of treatment in clinical studies [7]. Especially in surgical research and practice, daily assessments have been used to precisely describe the trajectory of symptom relief and functional recovery because the daily changes in symptoms in surgical patients have been found to be statistically significant [8-10]. However, whether the high frequency of assessment affects data quality has seldom been discussed in studies using longitudinal PROs as major outcomes.

What does “data quality” actually mean? Wang and Strong [11] and Kahn et al [12] proposed that data should be of sufficient quality to be of use to data consumers pursuing specific goals. For longitudinal data repositories, Weiskopf et al [13] characterized “data quality” as completeness. Charnock [14] conducted a systematic review in 2019 and reported that all papers referred to the importance of accuracy and completeness when evaluating data quality. Currently, data quality evaluations in longitudinal studies have focused on missing assessments [15-17]. However, other issues, such as item nonresponse and sample bias, have emerged over time [18,19], and these issues may impact data availability and consistent interpretations. Recent studies reported that repeated source data verification could improve accuracy and completeness by 40% [19], and better data quality could improve epidemiological inferences [20]. Additionally, partly due to the lack of an international definition of “error” [21], very few descriptions of the determinants of poor data quality have been provided in clinical studies [22]. Thus, there is an urgent need to characterize types of errors and the factors that affect longitudinal data quality to enable more interpretable results to be obtained from more complete data.

Paper-and-pencil (P&P) or electronic-based assessment of PRO (electronic PRO [ePRO]) are the 2 common modes used in clinical practice [23,24]. Compared with the P&P method, ePRO is more likely to generate complete data [17]; results in fewer data entry errors [25]; is more user friendly [26]; results in a shorter turnaround time [27]; and allows data to be processed, reviewed, and disseminated quickly [28,29]. Currently, interactive ePRO assessments can provide immediate feedback from patients [30] and are a convenient means of monitoring patients and delivering early warnings to clinicians [31,32]. In surgical research, due to the daily changes in symptoms after surgery [8-10], daily ePRO assessments have been used to precisely describe symptom relief and functional recovery. However, the often-mentioned disadvantages of ePRO assessments are sample bias [15,22,33] and a lower response rate [15-17,33]. Thus, generating a profile of the quality of data obtained with P&P and ePRO assessments will guide the appropriate selection of the mode of assessment.

Objectives

Daily PRO data collected via either P&P or ePRO assessments over the course of recovery from thoracic surgery for malignant or benign lung tumors were used in this analysis, with the following aims: (1) to describe error patterns in PRO data collected via the 2 major PRO measurement modes (ie, P&P and ePRO); (2) to identify factors influencing the incidence of errors; and (3) to generate profiles of the trajectories of errors over the course of a high-frequency data collection schedule.

Methods

Data Sources

Data were extracted from 2 prospective studies: 1 observational study [34] and 1 randomized controlled trial (RCT) [35]. The 2 original studies were approved by the Ethics Committee of Sichuan Cancer Hospital (No. SCCHEC-02-2017-042 and No. SCCHEC-02-2018-045).

All patients were assessed with the MD Anderson Symptom Inventory Lung Cancer Module (MDASI-LC) [36] and the single-item QOL scale [37] within 3 days before surgery and then daily after surgery until discharge or for up to 14 days if the patient stayed in the hospital for longer than 14 days after surgery. The MDASI-LC consists of 2 parts. Part I includes not only items regarding 13 core symptoms but also 3 items specific to lung cancer. Part II includes 6 interference items.

All data collection communications with medical staff were conducted face-to-face, and reminders were provided in the hospitals. Patients were asked to consider their symptoms over the previous 24 hours. When a participant completed and submitted a survey, he or she was not able to later modify the answers. On P&P assessments, signatures and data were collected from the patients and researchers for each record. Any time a patient modified a P&P form, the patient was asked to sign below the modified item. Assessment through ePRO only required the patient’s e-signature for each record.

The observational study used P&P, ePRO, phone-to-paper, and mixed assessments, while the RCT used only ePRO assessments. All PRO data were stored in the REDCap [38,39] online management system. EPRO data were automatically imported into REDCap within 24 hours, whereas the P&P forms were manually entered into this platform. Both studies were approved by the ethics committees of all participating hospitals. All participants signed informed consent forms [34,35].

For the P&P assessments, the original paper questionnaires were first checked by the data collectors for amendable errors (eg, missing researcher signatures at the end of completed

questionnaires). After both the P&P and ePRO data were entered into REDCap, the database was closed and sent for a data audit by a third team. The classification of errors was performed by 2 independent data management experts (QS and WD) with experience in clinical research data management. Inconsistencies were discussed within the audit team to reach a consensus. Data with errors identified during the audit were then entered into an electronic database in REDCap by 2 independent investigators (HY and QY) and cross-checked. The audit included (1) the withdrawal rate of each study; (2) patient compliance with the scheduled times of the assessments; (3) the completeness and accuracy of PRO forms with regard to individual items; and (4) rate of missing signatures and dates of completion.

Outcome Defined

Six types of errors were summarized into 2 groups, namely, incompleteness and inaccuracy, and used as indicators of PRO data quality:

- Incompleteness: any missing (1) individual items; (2) patient signatures; (3) researcher signatures; or (4) dates of completion.
- Inaccuracy: any (1) multiple selections for 1 item or (2) missing patient signature on any modified answer.

When any type of error mentioned above was found for any item on the MDASI-LC or QOL scale, it was counted as 1 error, and the corresponding patient was defined as a *patient with error*. A record of an error was defined as any error found for each PRO instrument (MDASI-LC or single-item QOL). A time point with any error in the record was labeled a *time point with an error*. Overall errors refer to all errors of all types in all records.

Data Analysis and Management

Reporting was performed according to STROBE guidelines [40]. To be included in the analysis, a patient must have provided PRO data at baseline and at least one additional time point during follow-up. We used the mean (SD) or median (range) for continuous variables and frequency (%) for categorical variables to describe the variables. Differences were analyzed using the 2-sample independent *t*-test, 2-sample Wilcoxon test, chi-square test, and Fisher exact test as appropriate. The withdrawal rate refers to the proportion of patients who did not provide a response to the assessment prior to the day of discharge. Patient compliance was calculated as the number of PRO assessments returned divided by the number of PRO assessments that should have been returned. We analyzed at most 8 time points (1 time before surgery and 7 days after surgery) when creating the profiles of the trajectories of the errors over time.

A multivariate generalized estimating equation (GEE) model was constructed to select and estimate the associations between potential risk factors and the incidence of errors for each mode of assessment. The factors included age (≤55 year vs. >55 year), sex (male vs. female), education (median school graduate or below vs. above), employment status (employed vs. other), surgical approach (video-assisted thoracoscopic surgery [VATS] vs. thoracotomy), hospital type (provincial vs. municipal or county level), BMI (≤23.9 kg/m^2 vs. >23.9 kg/m^2), smoking status (yes vs. no), Charlson Comorbidity Index (CCI) score (≤1 vs. >1), number of chest tubes (1 vs. 2), disease type (not-lung cancer vs. lung cancer with pathological tumor–node–metastasis [pTNM] stage>I, lung cancer with pTNM stage ≤I vs. lung cancer with pTNM stage>I), and postoperative hospital stay days (>6 vs. ≤6). The effect of risk factors is presented as odds ratios (ORs) with 95% CIs. Using Bonferroni correction [41] for multiple comparisons of risk factor identification, statistical significance level was set at the adjusted cutoff of $P<.004$, adjusted by the number of risk factors (0.05/number of risk factors).

The GEE model was also used to describe the trajectories of the incidences of errors over the 7 time points after surgery between those who used the P&P and ePRO assessments. The incidence of all errors or missing items was the dependent variable and the baseline covariates (the significant variables in the previous GEE model analysis), days after surgery (as a continuous variable), assessment modes, and the interaction between time and assessment mode were the independent variables. The binomial distribution and logit link function were adopted in all models. Co-variance structure types, such as unstructured, autoregressive, independent, exchangeable, and compound symmetric, were compared via quasi-likelihood under the independence model criterion (QIC). The models with QICs closest to 0 were closed as the final models. Two-piecewise random coefficient models were used to analyze trends before and after surgery. Time points with the highest proportion of errors were defined as the change points in the 2-piecewise models. All *P* values were 2 tailed, and statistical significance was set at the conventional cutoff of $P<.05$. All data analyses were performed using the statistical software SAS (version 9.4; SAS Institute).

Results

Participants

We extracted data pertaining to patients scheduled for lung surgery from the observational study (n=512) and the RCT (n=166). Thirty-six patients were excluded because they used phone-to-paper or mixed assessments, and 13 patients had only 1 PRO record. Finally, 629 patients responded to either P&P (n=440) or ePRO (n=189) assessments.

Patient characteristics are presented in Table 1. Compared with those using P&P assessments, patients using ePRO assessments were younger (51.5 vs. 55.5; $P<.001$), had higher levels of education (67.2% [127/189] vs. 50.0% [220/440]; $P<.001$), lower CCI scores (75.7% [143/189] vs. 60.7% [267/440]; $P<.001$), earlier stages of disease (compare with lung cancer with pTNM stage >I, 85.2% [161/189] vs. 68.0% [299/440]; $P<.001$), were more likely to have undergone VATS (93.1% [176/189] vs. 81.4% [358/440]; $P<.001$), and had shorter postoperative hospital stay (5 days vs. 6 days; $P<.001$). However, the differences in employment status, hospital type, and BMI were not significant (Table 1).

Table 1. Baseline characteristics of participants who filled out the P&P[a] or ePRO[b] assessment.[c]

Variables	ePRO (n=189)	P&P (n=440)	*P* value[d]
Age (years), mean (SD)	51.5 (10.8)	55.5 (10.3)	*<.001* [e]
Postoperative hospital stay (days), median (range)	5 (1-25)	6 (2-41)	*<.001* [f]
Gender, n (%)			*<.001* [g]
Male	73 (38.6)	247 (56.1)	
Female	116 (61.4)	193 (43.9)	
Education , n (%)			*<.001* [g]
Middle school or below	62 (32.8)	220 (50.0)	
Higher than middle school	127 (67.2)	220 (50.0)	
Employment status , n (%)			.62[g]
Employed	85 (45.0)	198 (45.0)	
Unemployed, peasant, retired, other	104 (55.0)	242 (55.0)	
Surgical approach , n (%)			*<.001* [g]
Video-assisted thoracoscopic surgery	176 (93.1)	358 (81.4)	
Thoracotomy	13 (6.9)	82 (18.6)	
Hospital type , n (%)			.89[g]
Provincial level	164 (86.8)	380 (86.4)	
Municipal or county level	25 (13.2)	60 (13.6)	
BMI (kg/m^2) , n (%)			.15[g]
≤23.9	130 (68.8)	276 (62.7)	
>23.9	59 (31.2)	168 (38.2)	
No smoking history[h], n (%)	151 (79.9)	262 (59.5)	*<.001* [g]
Charlson Comorbidity Index score , n (%)			*<.001* [g]
≤1	143 (75.7)	267 (60.7)	
>1	46 (24.3)	173 (39.3)	
Chest tube , n (%)			*<.001* [g]
1	96 (50.8)	315 (71.6)	
2	93 (49.2)	125 (28.4)	
Disease type , n (%)			*<.001* [g]
Nonlung cancer	16 (8.5)	82 (18.6)	
Lung cancer with pTNM[i] stage ≤I	145 (76.7)	217 (49.3)	
Lung cancer with pTNM stage >I	28 (14.8)	141 (32.0)	

[a]P&P: paper and pencil.

[b]ePRO: electronic PRO (patient-reported outcome).

[c]No data missing for demographic and clinical characteristic variables.

[d]Statistically significant values are italicized (*P*<.05).

[e]Independent *t*-test.

[f]Wilcoxon 2-sample test.

[g]Chi-square test.

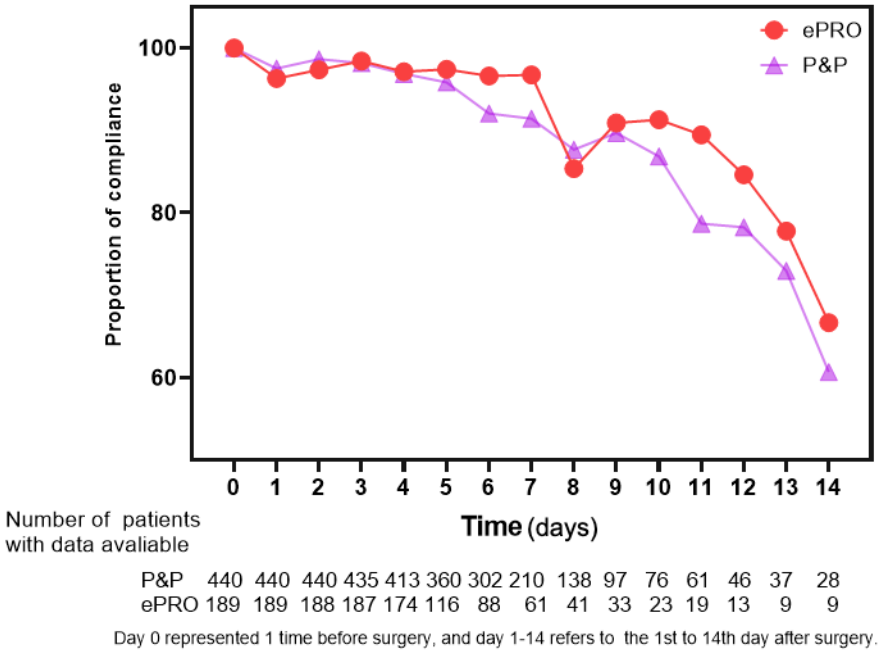
[h]Former or current smoker except no smoking history.

[i]pTNM: pathological tumor–node–metastasis.

Compliance With Scheduled Assessments Over Time

Of the 629 patients included in the analysis, 6.4% (28/440) of the patients in the P&P group and 3.7% (7/189) of the patients in the ePRO group withdrew from the studies during hospitalization. A total of 440 P&P patients generated 3347 PRO records, whereas 189 ePRO patients generated 1291 records. The compliance rates ranged from 67% (6/9 in POD 14) to 100% (189/189 before surgery) for the ePRO group and from 61% (17/28 in POD 14) to 100% (440/440 before surgery) for the P&P group over time (Figure 1).

Figure 1. Proportion of patient compliance during 15 timepoints between different modes. ePRO: electronic PRO, P&P: paper and pencil, PRO: patient-reported outcome.



Error Patterns

We found that 49.4% (311/629) of the patients had at least one error, and a total of 1654 errors were identified. In Multimedia Appendix 1, missing items (64.69%, 1070/1654) and modifications without signatures (27.99%, 463/1654) were the top 2 most frequently observed errors, followed by multiple selections for 1 item (3.02%, 50/1654), missing patient signatures (2.54%, 42/1654), missing researcher signatures (1.45%, 24/1654), and missing completion dates (0.30%, 5/1654).

Multiple selections for a single item, modifications without patient or researcher signatures, and missing completion dates were only identified on P&P assessments, accounting for 32.77% (542/1654). Shown in Table 2, significant differences in the number of involved patients were found for the overall errors (ePRO: 30.2% [57/189] vs. P&P: 57.7% [254/440]; $P<.001$) and missing items (ePRO: 28.6% [54/189] vs. P&P: 55.0% [242/440]; $P<.001$). Very few "missing patient signature" errors were identified, and the proportion did not differ between the ePRO and P&P groups (2.1% [4/189] vs. 1.8% [8/440]; $P=.76$).

The error rates of each item (including missing items, modifications without signatures, and multiple selections for 1 item) within PRO instruments are presented in Multimedia Appendix 2. Overall errors and missing items were found in 4% of the items pertaining to distress and interferes (mood and relations) on both types of assessments (ePRO and P&P).

Table 2. Counts and proportion of involved patients and errors between assessment modes (ePRO[a] vs. P&P[b]).

Error types	Errors (count), n		Involved patients, n (%)		*P* value[c]
	ePRO (n=189)	Paper (n=440)	ePRO (n=189)	Paper (n=440)	
Missing items	152	918	54 (28.6)	242 (55.0)	*<.001* [d]
Modifications without signatures	0	463	0 (0)	140 (31.8)	
Multiple selection for 1 item	0	50	0 (0)	42 (9.5)	
Missing patient signatures	14	28	4 (2.1)	8 (1.8)	.76[e]
Missing researcher signatures	0	24	0 (0)	11 (2.5)	
Missing completion dates	0	5	0 (0)	3 (0.7)	
Overall errors	166	1488	57 (30.2)	254 (57.7)	*<.001* [d]

[a]ePRO: electronic PRO (patient-reported outcome).

[b]P&P: paper and pencil.

[c]Statistically significant values are italicized (*P*<.05).

[d]Chi-square test.

[e]Fisher exact test.

Factors Contributing to the Incidence of Errors

As shown in Table 3, patients with lower education levels (OR 1.82, 95% CI 1.22-2.72; *P*=.003), those treated at provincial hospitals (OR 4.73, 95% CI 2.18-10.25; *P*<.001), and those with severe disease (lung cancer with pTNM stage >I vs. nonlung cancer: OR 2.70, 95% CI 1.53-4.75; *P*<.001) were more likely to generate errors in the ePRO group. In the P&P group, a lower level of education (OR 1.39, 95% CI 1.20-1.62; *P*<.001), treatment in a provincial hospital (OR 3.34, 95% CI 2.10-5.33; *P*<.001), severe disease (lung cancer with pTNM stage >I vs. nonlung cancer: OR 1.63, 95% CI 1.33-1.99; *P*<.001), being younger (OR 1.47, 95% CI 1.15-1.88; *P*=.002), male sex (OR 1.41, 95% CI 1.12-1.78; *P*=.003), thoracotomy (OR 1.28, 95% CI 1.13-1.46; *P*<.001), a higher CCI score (OR 1.58, 95% CI 1.36-1.84; *P*<.001), and more chest tubes (OR 1.66, 95% CI 1.26-2.17; *P*<.001) were associated with a higher risk of errors. The details of risk factors for missing items in P&P and ePRO are shown in Multimedia Appendix 3.

Table 3. Factors associated with the error incidence rate of participants who filled out the (ePRO[a] vs. P&P[b]) assessments[c].

Factors	ePRO (n=189)		Paper-and-pencil mode (n=440)	
	OR[d] (95% CI)	*P* value[e]	OR (95% CI)	*P* value[e]
Age (under 55 years vs. 55 years or older)	0.96 (0.48-1.93)	.91	1.47 (1.15-1.88)	*.002*
Gender (male vs. female)	0.93 (0.60-1.42)	.73	1.41 (1.12-1.78)	*.003*
Education (middle school or below vs. higher than middle school)	1.82 (1.22-2.72)	*.003*	1.39 (1.20-1.62)	*<.001*
Employment status (others vs. employed)	0.93 (0.65-1.34)	.71	1.15 (1.02-1.31)	.03
Surgical approach (thoracotomy vs. video-assisted thoracoscopic surgery)	1.95 (1.17-3.25)	.01	1.28 (1.13-1.46)	*<.001*
Hospital type (provincial level vs. municipal or county level)	4.73 (2.18-10.25)	*<.001*	3.34 (2.10-5.33)	*<.001*
BMI (>23.9 kg/m^2 vs. ≤23.9 kg/m^2)	1.36 (0.87-2.12)	.18	0.93 (0.79-1.10)	.40
Smoking status[f] (yes vs. no)	0.70 (0.47-1.03)	.07	1.14 (0.90-1.46)	.28
Charlson Comorbidity Index score (>1 vs. ≤1)	2.40 (1.11-5.20)	.03	1.58 (1.36-1.84)	*<.001*
Chest tube (2 vs. 1)	0.57 (0.37-0.89)	.01	1.66 (1.26-2.17)	*<.001*
Disease type				
Lung cancer with pTNM[g] stage ≤I vs. nonlung cancer	1.21 (0.85-1.72)	.29	1.17 (0.88-1.57)	.28
Lung cancer with pTNM stage >I vs. nonlung cancer	2.70 (1.53-4.75)	*<.001*	1.63 (1.33-1.99)	*<.001*
Postoperative hospital stay (6 days or above vs. under 6 days)	0.90 (0.56-1.47)	.69	1.12 (0.95-1.32)	.18

[a]ePRO: electronic PRO (patient-reported outcome).

[b]P&P: paper and pencil.

[c]Administration: generalized estimated equation model; $\alpha' = \alpha/12 = 0.0042$.

[d]OR: odds ratio.

[e]Statistically significant values are italicized ($P<.05$).

[f]Former or current smoker except no smoking.

[g]pTNM: pathological tumor–node–metastasis.

Trajectories of Errors

The trajectories of overall errors and missing items over time are illustrated for the ePRO and P&P assessments separately (Figure 2). In the P&P group, 14.8% (65/440) of patients made errors before surgery and then peaked on postoperative day 1 (POD 1; 117/440, 26.6%). The trajectory gradually decreased after surgery, but remained higher than that before surgery (17.2% [33/192] on POD 7). In the ePRO group, overall error was 3.2% (6/189) before surgery, followed by a continuous increase after surgery, peaking on POD 2 (13.1% [24/183]), and then gradually decreased but remained higher than that before surgery (POD 7 in 5.1% [3/59]). Similarly, missing items peaked on POD 1 in the P&P group (25.9% [111/429]) and on POD 2 in the ePRO group (12.0% [22/183]; Figure 2B). The details are presented in Table 4.

Figure 2. Error incidence rate of responded records with overall errors or item missing during 8 timepoints in hospital. ePRO: electronic PRO, P&P: paper and pencil, PRO: patient-reported outcome.

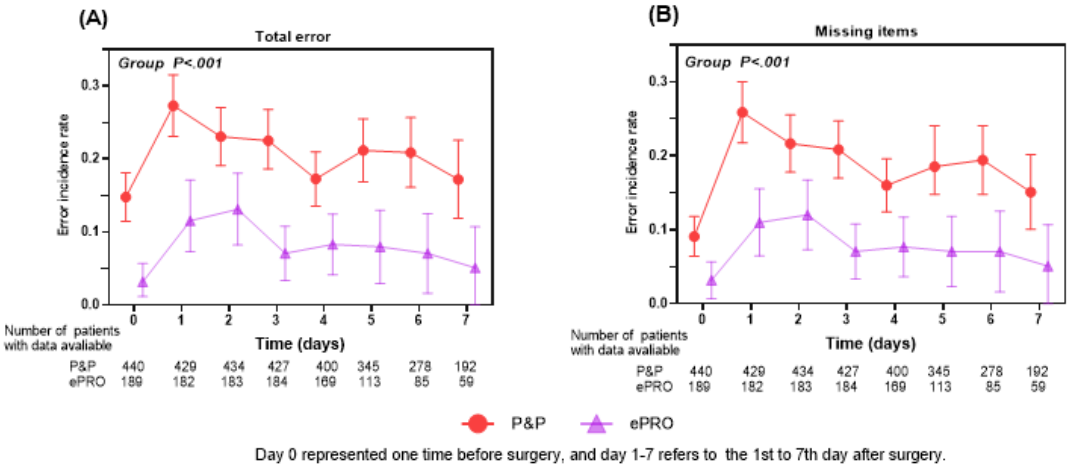


Table 4. Error incidence records of responded records with overall errors or item missing during 8 time points in hospital.[a]

Time (days)	Overall errors			Item missing			Completed patients	
	ePRO[b] (n=189), n (%)	P&P[c] (n=440), n (%)	*P* value[d,e]	ePRO[b] (n=189), n (%)	P&P[c] (n=440), n (%)	*P* value[d,e]	ePRO, n	P&P, n
0[f]	6/189 (3.2)	65/440 (14.8)	Mode=*.005*; time=*.03*; MT[g]=.69	6/189 (3.2)	40/440 (9.1)	Mode=*.005;* time=.06; MT[g]=.88	189	440
1	21/182 (11.5)	117/429 (27.3)		20/182 (11.0)	111/429 (25.9)		182	429
2	24/183 (13.1)	100/434 (23.0)		22/183 (12.0)	94/434 (21.7)		183	434
3	13/184 (7.1)	96/427 (22.5)		13/184 (7.1)	89/427 (20.8)		184	427
4	14/169 (8.3)	69/400 (17.3)		13/169 (7.7)	64/400 (16.0)		169	400
5	9/113 (8.0)	73/345 (21.2)		8/113 (7.1)	64/345 (18.6)		113	345
6	6/85 (7.1)	58/278 (20.9)		6/85 (7.1)	54/278 (19.4)		85	278
7	3/59 (5.1)	33/192 (17.2)		3/59 (5.1)	29/192 (15.1)		59	192

[a]Administration: generalized estimated equation (GEE) model.

[b]ePRO: electronic PRO (patient-reported outcome) (web-based).

[c]P&P: paper-and-pencil.

[d]Adjusted GEE model *P* values reported for time effect (as continual variable), mode effect (reference as P&P mode), interaction between mode and time effect (MT). All others are baseline covariant.

[e]Statistically significant values are italicized (*P*<.05).

[f]Day 0 represented the 1 time before surgery, and 1-7 refers to the 1st to day 7th after surgery.

[g]MT: interaction between mode effect and time effect.

The inflection time points were POD 2 for the ePRO assessment and POD 1 for the P&P assessment (Table 5). The incidence of errors on the ePRO assessments significantly increased from before surgery to POD 2 (estimate=0.51; *P*=.01, in model 2) and significantly decreased after POD 2 (estimate=–0.21; *P*<.001). However, errors on the P&P assessment significantly increased over the first 2 assessment time points (estimate=0.73; *P*<.001, in model 2) and slightly decreased after POD 1 (estimate=–0.10; *P*<.001). The details of item missing using 2-piecewise model are described in Multimedia Appendix 4.

Table 5. Two-piecewise regression analysis for each mode with overall errors during 8 time points in hospital.[a]

Mode	Overall errors			
	Estimate 1[b] (standard error)	*P* value[c]	Estimate 2[d] (standard error)	*P* value[c]
Electronic PRO mode				
Model 1[e]	0.50 (0.20)	*.01*	–0.20 (0.04)	*<.001*
Model 2[f]	0.51 (0.20)	*.01*	–0.21 (0.05)	*<.001*
Model 3[g]	0.55 (0.19)	*.004*	–0.24 (0.05)	*<.001*
Paper and pencil mode				
Model 4[e]	0.67 (0.06)	*<.001*	–0.08 (0.02)	*<.001*
Model 5[h]	0.73 (0.05)	*<.001*	–0.10 (0.02)	*<.001*
Model 6[g]	0.74 (0.05)	*<.001*	–0.11 (0.02)	*<.001*

[a]Administration: 2-piecewise model; inflection point, POD (postoperative day) 1 for P&P (paper and pencil) and POD 2 for ePRO (electronic PRO [patient-reported outcome]) .

[b]Estimate 1: piecewise regression coefficient on the left side of the inflection point, from before surgery to POD 2 in the ePRO mode or from before surgery to POD 1 in the P&P mode after surgery

[c]Statistically significant values are italicized ($P<.05$).

[d]Estimate 2: piecewise regression coefficient on the right side of the inflection point.

[e]Models 1 and 4: no adjustment.

[f]Model 2: adjustment for education, hospital level, and disease type.

[g]Models 3 and 6: adjustment for age group, gender, education, employment, surgical approach, hospital type, BMI, smoking history, Charlson Comorbidity Index score, chest tube, disease type, and postoperative hospital stay (days).

[h]Model 5: adjustment for age group, gender, education, surgical approach, hospital type, Charlson Comorbidity Index score, chest tube, and disease type.

Discussion

Principal Findings

For the first time, using data from studies that included PROs as major outcomes in the setting of thoracic surgery, we profiled 6 types and 2 trajectories of errors for PRO data collected daily using 2 major assessments (ePRO or P&P). Nearly one-fifth of the records and half of the patients had errors when longitudinal PROs were used as outcomes, even when a quality check was implemented immediately after the completion of data collection. We demonstrated that, compared with the P&P assessment, the ePRO assessment had higher compliance, which is necessary to maintain data quality, but needed more time for patient adaptation. In addition, significant selection bias was identified for the ePRO assessment, with younger, better educated, and more physically active patients being more likely to use. This quantification of the quality of frequently collected PRO data might support study design, data quality control, and data audits for surgical studies using PROs as outcomes and will help guide resource allocation when implementing PRO-based surgical patient care.

Magnitude of Data Errors

The ePRO assessment had fewer errors. Over one-third of the errors occurred on P&P assessments, and these were errors that could be avoided by using the ePRO assessment. One study described missing items on anxiety questionnaires at 3 assessment points, and the results were as follows: 31.8% for P&P versus 2.08% for ePRO in the hospital [15]. Another study that investigated food-frequency questionnaires at 2 time points over 10 years revealed that the average rate of missing items on the form was 9% for P&P assessments and 3% for the electronic version [17]. The lower rates of errors observed in those 2 studies may be attributed to the lower frequency of measurement and the younger participants. Zeleke et al [42] analyzed 2492 records in an RCT involving healthy people and reported that 41.89% of the paper records and 30.89% of the electronic records had 1 or more types of data quality issues. Compared with those studies, our analysis, which had clear definitions of data inaccuracy and incompleteness, suggested the need for careful data quality monitoring plans in studies that require frequent assessments of PROs.

Missing items on assessments of PROs is a core issue and is nearly ubiquitous in clinical research. In this study, missing items accounted for a significant proportion (over three-fifths) of all errors. There is strong evidence that much of these missed items occur at random and are therefore almost impossible to eliminate in the real world [19,43,44]. Our results showed that missing items decreased by one-fifth when ePRO assessment was used, indicating that using this format could improve PRO data quality in further studies.

Adaptation

The trajectories in errors significantly changed each day during the perioperative period, and different trends were observed for each assessment mode. Interestingly, constant trends, with an initial increase followed by a decrease over time, were observed with both the P&P and ePRO assessments in this study, whereas

the results in a similar study showed random peaks and irregular trends when the data were presented according to the date of collection [42]. There are 2 possible reasons for the difference. First, the sequence of time points that this study followed merely ordered the data according to the natural progression of days, from day 1 of the survey to day 25 of the survey, whereas our analysis considered the sequence of response time points for each patient. Second, that study was performed at public health and demographic surveillance sites, whereas we targeted surgical patients in hospitals. By contrast, a learning curve usually occurs for the use of a new technological progress (reflected by a decreasing error rate) as a function of the accumulation of experience over time [45]. Errors peaked on POD 1 for P&P assessments and on POD 2 for ePRO assessments, suggesting that patients took less time to adapt to the former. Studies have reported that more experience and time are needed to adapt to electronic methods [46]. Basch et al [4] found that patients with prior computer use experience benefited relatively more from the web-based PRO monitoring and alerting system.

In general, paper-based assessments are expected to be the first choice [17]. P&P is still a major method of assessment in clinical research, especially for older, poorer, or sicker patients. To accommodate a more representative patient set, ePRO needs to be made more user-friendly. For example, reducing the complexity of operating the interface, adding or optimizing automated interactive voice functions, and designing automated telephone systems outside of the hospital should be considered [47]. Given the convincing equality in measuring patient perception, a mixed-mode system involving both P&P and ePRO assessments could be a better choice. The preferred option might be ePRO assessments, with P&P assessments as the secondary choice for almost all patients in clinical studies.

What Are the Factors That Influenced Data Quality?

In this analysis, patients treated in provincial hospitals were more likely to produce poor-quality PRO data regardless of whether they used the P&P or ePRO assessments. The explanation was that the majority of patients and heaviest clinical workload are concentrated in provincial hospitals in China [48]. Medical staff in provincial hospitals are busier than those in municipal or county-level hospitals in routine clinical practice, which may result in less effort given to data monitoring. For any patient-centered practice or research, more efforts are required to obtain better data availability and accuracy in health care system. Other shared factors affecting errors in both modes are education level and physical status. Therefore, we suggest that there should be prespecified means of assistance provided to participants who are more likely to struggle to complete the assessments [49]. For example, measures might be taken to help patients complete scheduled PRO assessments when they have greater difficulties filling in the form [50]. Compared with P&P assessments, ePRO assessments had fewer risk factors for poor data quality. One possible explanation might be the homogeneous population using ePRO assessments due to biased sampling, as their use requires a certain level of education [51].

Limitations

We acknowledge that the results are limited by the potential sample bias and the differences in study designs and data collection tools. We may have overestimated the differences between the P&P and ePRO assessments because RCTs are managed better than observational studies [52], although the same team of clinical coordinators and same data quality control standard operating procedure were used for both projects. A second limitation is that the data were only collected during hospitalization because almost all ePRO assessments were administered after discharge in our study. This is similar to a study that showed that the ePRO assessment was more cost-effective and user friendly for clinical staff and patients [16,31] and suggested that there is a trend in the implementation of ePRO assessments in clinical research. Finally, this study lacks evidence of the equivalence of the data collected with the 2 forms of assessment, and therefore cannot state whether the data collected with the 2 assessments are equally valid. Further research is needed to confirm these results.

Conclusions

In conclusion, this study with substantial sample and longitudinal design demonstrates the pros and cons of the 2 most commonly used methods (ePRO and P&P), which will help promote web-based patient care [53]. It is possible to improve the quality of longitudinal PRO data by using web-based assessments. Although ePRO was found to be superior to P&P in terms of data quality, ePRO-related sampling bias should be taken into consideration when designing clinical research using longitudinal PROs as a major outcome.

Alternatively, providing the option of using either the ePRO or the P&P assessment would improve the representativeness of samples if the comparativeness of the data obtained with the ePRO and P&P assessments is confirmed by well-designed equivalence studies.

Acknowledgments

This work was supported by the Chongqing Graduate Student Research Innovation Project (CYS20209), National Natural Science Foundation of China (No. 81872506), and Sichuan Science and Technology Program (No. 2019YFH0070). The authors thank Yaqing Wang, Jia Liao, Xiaozun Yang, and Shaohua Xie (Department of Thoracic Surgery, Sichuan Cancer Hospital), Wenhong Feng (Department of Thoracic and Cardiovascular Surgery, Jiangyou People's Hospital), Yuanqiang Zhang (Department of Cardiothoracic Surgery, Zigong First People's Hospital), Yunfei Mu (Department of Thoracic Surgery, Chengdu Third People's Hospital), Rui Zhang (Department of Thoracic Surgery, Chengdu Seventh People's Hospital), and Xiaoqing Liao (Department of Cardiothoracic Surgical Oncology, Dazhu County People's Hospital, Dazhu County) for their contributions to the data collection of this paper. We also thank our senior fellow apprentice, Li Tang, for statistical method support.

Authors' Contributions

All authors took part in the study concept and design; acquisition, analysis, and interpretation of data and in the final approval of the version to be published. HY, QS, and XW were responsible for drafting of the abstract. QS, WD, and XW contributed to revising the article critically for important intellectual content. HY, QY, YN, WX, and YP performed statistical analysis. QS and WD obtained funding and offered administrative, technical, or material support. The study was supervised by QS. Both HY and QS had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the error pattern.
[PNG File , 147 KB - jmir_v23i11e28915_app1.png]

Multimedia Appendix 2
Item missing and overall errors within each item of MD Anderson Symptom Inventory Lung Cancer Module (MDASI-LC) and quality of life (QOL).
[DOC File , 67 KB - jmir_v23i11e28915_app2.doc]

Multimedia Appendix 3
Factors associated with the item missing incidence rate, of participants who filled out the ePRO (ePRO: electronic PRO [patient-reported outcome]), P&P (paper and pencil) and overall modes.
[DOC File , 50 KB - jmir_v23i11e28915_app3.doc]

Multimedia Appendix 4
Two-piecewise regression analysis for each mode with item missing during 8 days in hospital.
[DOC File , 37 KB - jmir_v23i11e28915_app4.doc]

References

1. Wei X, Yu H, Dai W, Mu Y, Wang Y, Liao J, et al. Patient-reported outcomes of video-assisted thoracoscopic surgery versus thoracotomy for locally advanced lung cancer: a longitudinal cohort study. Ann Surg Oncol 2021 Apr 20:e1 (forthcoming). [doi: 10.1245/s10434-021-09981-1] [Medline: 33880671]
2. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, the SPIRIT-PRO Group, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA 2018 Feb 06;319(5):483-494. [doi: 10.1001/jama.2017.21903] [Medline: 29411037]
3. Calvert M, Blazeby J, Altman D, Revicki D, Moher D, Brundage M, et al. Reporting of patient-reported outcomes in randomized trials. JAMA 2013 Feb 27;309(8):814-N/A. [doi: 10.1001/jama.2013.879]
4. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. JCO 2016 Feb 20;34(6):557-565. [doi: 10.1200/jco.2015.63.0830]
5. Denis F, Basch E, Lethrosne C, Pourel N, Molinier O, Pointreau Y, et al. Randomized trial comparing a web-mediated follow-up via patient-reported outcomes (PRO) vs. routine surveillance in lung cancer patients: final results. JCO 2018 May 20;36(15_suppl):6500-6500. [doi: 10.1200/jco.2018.36.15_suppl.6500]
6. Bendixen M, Jørgensen O, Kronborg C, Andersen C, Licht P. Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. The Lancet Oncology 2016 Jun;17(6):836-844. [doi: 10.1016/S1470-2045(16)00173-X]
7. Aiyegbusi OL, Kyte D, Cockwell P, Anderson N, Calvert M. A patient-centred approach to measuring quality in kidney care. Current Opinion in Nephrology and Hypertension 2017;26(6):442-449. [doi: 10.1097/mnh.0000000000000357]
8. Meyer LA, Nick AM, Shi Q, Wang XS, Williams L, Brock T, et al. Perioperative trajectory of patient reported symptoms: a pilot study in gynecologic oncology patients. Gynecol Oncol 2015 Mar;136(3):440-445 [FREE Full text] [doi: 10.1016/j.ygyno.2015.01.547] [Medline: 25647455]
9. Shi Q, Mendoza TR, Wang XS, Cleeland CS. Using a symptom-specific instrument to measure patient-reported daily functioning in patients with cancer. Eur J Cancer 2016 Nov;67:83-90. [doi: 10.1016/j.ejca.2016.07.027] [Medline: 27620946]
10. Meyer LA, Shi Q, Lasala J, Iniesta MD, Lin HK, Nick AM, et al. Comparison of patient reported symptom burden on an enhanced recovery after surgery (ERAS) care pathway in patients with ovarian cancer undergoing primary vs. interval

tumor reductive surgery. Gynecol Oncol 2019 Mar;152(3):501-508 [FREE Full text] [doi: 10.1016/j.ygyno.2018.10.044] [Medline: 30876495]
11. Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
12. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care 2012 Jul;50 Suppl:S21-S29 [FREE Full text] [doi: 10.1097/MLR.0b013e318257dd67] [Medline: 22692254]
13. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform 2013 Oct;46(5):830-836. [doi: 10.1016/j.jbi.2013.06.010] [Medline: 23820016]
14. Charnock V. Electronic healthcare records and data quality. Health Info Libr J 2019 Mar;36(1):91-95. [doi: 10.1111/hir.12249] [Medline: 30811882]
15. Barentsz MW, Wessels H, van Diest PJ, Pijnappel RM, Haaring C, van der Pol CC, et al. Tablet, web-based, or paper questionnaires for measuring anxiety in patients suspected of breast cancer: patients' preferences and quality of collected data. J Med Internet Res 2014 Oct 31;16(10):e239 [FREE Full text] [doi: 10.2196/jmir.3578] [Medline: 25364951]
16. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. J Med Internet Res 2018 Jan 23;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353] [Medline: 29362206]
17. Zazpe I, Santiago S, De la Fuente-Arrillaga C, Nuñez-Córdoba JM, Bes-Rastrollo M, Martínez-González MA. Paper-based versus web-based versions of self-administered questionnaires, including food-frequency questionnaires: prospective cohort study. JMIR Public Health Surveill 2019 Oct 01;5(4):e11997 [FREE Full text] [doi: 10.2196/11997] [Medline: 31573924]
18. Rübsamen N, Akmatov MK, Castell S, Karch A, Mikolajczyk RT. Comparison of response patterns in different survey designs: a longitudinal panel with mixed-mode and online-only design. Emerg Themes Epidemiol 2017;14:4 [FREE Full text] [doi: 10.1186/s12982-017-0058-2] [Medline: 28344629]
19. Houston L, Probst Y, Martin A. Assessing data quality and the variability of source data verification auditing methods in clinical research settings. J Biomed Inform 2018 Jul;83:25-32 [FREE Full text] [doi: 10.1016/j.jbi.2018.05.010] [Medline: 29783038]
20. Giganti MJ, Shepherd BE, Caro-Vega Y, Luz PM, Rebeiro PF, Maia M, et al. The impact of data quality and source data verification on epidemiologic inference: a practical application using HIV observational data. BMC Public Health 2019 Dec 30;19(1):1748 [FREE Full text] [doi: 10.1186/s12889-019-8105-2] [Medline: 31888571]
21. Houston L, Probst Y, Humphries A. Measuring data quality through a source data verification audit in a clinical research setting. Stud Health Technol Inform 2015;214:107-113. [Medline: 26210426]
22. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042 [FREE Full text] [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
23. Gregory KE, Radovinsky L. Research strategies that result in optimal data collection from the patient medical record. Appl Nurs Res 2012 May;25(2):108-116 [FREE Full text] [doi: 10.1016/j.apnr.2010.02.004] [Medline: 20974093]
24. King JD, Buolamwini J, Cromwell EA, Panfel A, Teferi T, Zerihun M, et al. A novel electronic data collection system for large-scale surveys of neglected tropical diseases. PLoS One 2013;8(9):e74570 [FREE Full text] [doi: 10.1371/journal.pone.0074570] [Medline: 24066147]
25. van Gelder MMHJ, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 01;172(11):1292-1298. [doi: 10.1093/aje/kwq291] [Medline: 20880962]
26. Beebe TJ, Jacobson RM, Jenkins SM, Lackore KA, Rutten LJF. Testing the impact of mixed-mode designs (mail and web) and multiple contact attempts within mode (mail or web) on clinician survey response. Health Serv Res 2018 Aug;53 Suppl 1:3070-3083 [FREE Full text] [doi: 10.1111/1475-6773.12827] [Medline: 29355920]
27. Smith AB, King M, Butow P, Olver I. A comparison of data quality and practicality of online versus postal questionnaires in a sample of testicular cancer survivors. Psychooncology 2013 Jan;22(1):233-237. [doi: 10.1002/pon.2052] [Medline: 21954102]
28. Dickinson FM, McCauley M, Madaj B, van den Broek N. Using electronic tablets for data collection for healthcare service and maternal health assessments in low resource settings: lessons learnt. BMC Health Serv Res 2019 May 27;19(1):336 [FREE Full text] [doi: 10.1186/s12913-019-4161-7] [Medline: 31133032]
29. Yang NS, Ward BW, Cummings NA. Patient health information shared electronically by office-based physicians: United States, 2015. Natl Health Stat Report 2018 Aug(115):1-9 [FREE Full text] [Medline: 30248005]
30. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333 [FREE Full text] [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
31. Kaur M, Pusic A, Gibbons C, Klassen AF. Implementing electronic patient-reported outcome measures in outpatient cosmetic surgery clinics: an exploratory qualitative study. Aesthet Surg J 2019 May 16;39(6):687-695. [doi: 10.1093/asj/sjy280] [Medline: 30335134]

32. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Dutton M, Walmsley-Allen N, et al. Patient and clinician perspectives on electronic patient-reported outcome measures in the management of advanced CKD: a qualitative study. Am J Kidney Dis 2019 Aug;74(2):167-178. [doi: 10.1053/j.ajkd.2019.02.011] [Medline: 31003865]
33. Horevoorts NJ, Vissers PA, Mols F, Thong MS, van de Poll-Franse LV. Response rates for patient-reported outcomes using web-based versus paper questionnaires: comparison of two invitational methods in older colorectal cancer patients. J Med Internet Res 2015 May 07;17(5):e111 [FREE Full text] [doi: 10.2196/jmir.3741] [Medline: 25953059]
34. Dai W, Xie S, Zhang R, Wei X, Wu C, Zhang Y, et al. Developing and validating utility parameters to establish patient-reported outcome-based perioperative symptom management in patients with lung cancer: a multicentre, prospective, observational cohort study protocol. BMJ Open 2019 Oct 28;9(10):e030726 [FREE Full text] [doi: 10.1136/bmjopen-2019-030726] [Medline: 31662377]
35. Dai W, Zhang Y, Feng W, Liao X, Mu Y, Zhang R, et al. Using patient-reported outcomes to manage postoperative symptoms in patients with lung cancer: protocol for a multicentre, randomised controlled trial. BMJ Open 2019 Aug 26;9(8):e030041 [FREE Full text] [doi: 10.1136/bmjopen-2019-030041] [Medline: 31455710]
36. Mendoza TR, Wang XS, Lu C, Palos GR, Liao Z, Mobley GM, et al. Measuring the symptom burden of lung cancer: the validity and utility of the lung cancer module of the M. D. Anderson Symptom Inventory. Oncologist 2011;16(2):217-227 [FREE Full text] [doi: 10.1634/theoncologist.2010-0193] [Medline: 21285393]
37. Sloan JA, Loprinzi CL, Kuross SA, Miser AW, O'Fallon JR, Mahoney MR, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. JCO 1998 Nov;16(11):3662-3673. [doi: 10.1200/jco.1998.16.11.3662]
38. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
40. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet 2007 Oct;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X]
41. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170 [FREE Full text] [doi: 10.1136/bmj.310.6973.170] [Medline: 7833759]
42. Zeleke AA, Worku AG, Demissie A, Otto-Sobotka F, Wilken M, Lipprandt M, et al. Evaluation of electronic and paper-pen data capturing tools for data quality in a public health survey in a health and demographic surveillance site, Ethiopia: randomized controlled crossover health care information technology evaluation. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10995 [FREE Full text] [doi: 10.2196/10995] [Medline: 30741642]
43. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. J Natl Cancer Inst 2017 Nov 01;109(11):djx187. [doi: 10.1093/jnci/djx187] [Medline: 29059439]
44. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
45. Campbell RD, Hecker KG, Biau DJ, Pang DSJ. Student attainment of proficiency in a clinical skill: the assessment of individual learning curves. PLoS One 2014;9(2):e88526 [FREE Full text] [doi: 10.1371/journal.pone.0088526] [Medline: 24586337]
46. Larcher A, De Naeyer G, Turri F, Dell'Oglio P, Capitanio U, Collins JW, ERUS Educational Working Groupthe Young Academic Urologist Working Group on Robot-assisted Surgery. The ERUS Curriculum for robot-assisted partial nephrectomy: structure definition and pilot clinical validation. Eur Urol 2019 Jun;75(6):1023-1031. [doi: 10.1016/j.eururo.2019.02.031] [Medline: 30979635]
47. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol 2011 Mar 10;29(8):994-1000 [FREE Full text] [doi: 10.1200/JCO.2010.29.8315] [Medline: 21282546]
48. Fang P, Li W. Analysis of the leading role of tertiary public hospitals in medical alliance. Hospital management in China 2018;38(05):1-3.
49. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res 2014 Oct;23(5):440-459. [doi: 10.1177/0962280213476378] [Medline: 23427225]
50. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. BMJ Open 2016 Jun 15;6(6):e010938 [FREE Full text] [doi: 10.1136/bmjopen-2015-010938] [Medline: 27311907]
51. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]

52. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med 2011 May;8(5):e1001026 [FREE Full text] [doi: 10.1371/journal.pmed.1001026] [Medline: 21559325]
53. Nagy G, Vári SG, Mezo T, Bogár L, Fülesdi B. Hungarian web-based nationwide anaesthesia and intensive care data collection and reporting system: its development and experience from the first 5 yr. Br J Anaesth 2010 Jun;104(6):711-716 [FREE Full text] [doi: 10.1093/bja/aeq085] [Medline: 20421226]

Abbreviations

CCI: Charlson Comorbidity Index
ePRO: electronic PRO
GEE: generalized estimating equation
MDASI-LC: MD Anderson Symptom Inventory Lung Cancer Module
OR: odds ratio
P&P: paper and pencil
POD: postoperative day
PRO: patient-reported outcome
pTNM: pathological tumor–node–metastasis
QIC: quasi-likelihood under the independence model criterion
QOL: quality of life
RCT: randomized controlled trial
VATS: video-assisted thoracoscopic surgery

Edited by R Kukafka, G Eysenbach; submitted 18.03.21; peer-reviewed by I Adeleke; comments to author 14.05.21; revised version received 21.05.21; accepted 03.10.21; published 09.11.21.

Please cite as:
Yu H, Yu Q, Nie Y, Xu W, Pu Y, Dai W, Wei X, Shi Q
Data Quality of Longitudinally Collected Patient-Reported Outcomes After Thoracic Surgery: Comparison of Paper- and Web-Based Assessments
J Med Internet Res 2021;23(11):e28915
URL: https://www.jmir.org/2021/11/e28915
doi:10.2196/28915
PMID:34751657

Review

Effects of Digital Technologies on Older People's Access to Health and Social Care: Umbrella Review

Tafadzwa Patience Kunonga[1*], MPH; Gemma Frances Spiers[1*], PhD; Fiona R Beyer[1*], PGD; Barbara Hanratty[1*], PhD; Elisabeth Boulton[2*], PhD; Alex Hall[2*], PhD; Peter Bower[2*], PhD; Chris Todd[2*], PhD; Dawn Craig[1*], MSc

[1]National Institute for Health Research Older People and Frailty Policy Research Unit, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]National Institute for Health Research Older People and Frailty Policy Research Unit, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Tafadzwa Patience Kunonga, MPH
National Institute for Health Research Older People and Frailty Policy Research Unit
Population Health Sciences Institute
Newcastle University
Newcastle Biomedical Research Building, Campus for Ageing and Vitality
Westgate Road
Newcastle upon Tyne, NE4 5PL
United Kingdom
Phone: 44 1912086260
Email: patience.kunonga@newcastle.ac.uk

Abstract

Background: The 2020 COVID-19 pandemic prompted the rapid implementation of new and existing digital technologies to facilitate access to health and care services during physical distancing. Older people may be disadvantaged in that regard if they are unable to use or have access to smartphones, tablets, computers, or other technologies.

Objective: In this study, we synthesized evidence on the impact of digital technologies on older adults' access to health and social services.

Methods: We conducted an umbrella review of systematic reviews published from January 2000 to October 2019 using comprehensive searches of 6 databases. We looked for reviews in a population of adults aged ≥65 years in any setting, reporting outcomes related to the impact of technologies on access to health and social care services.

Results: A total of 7 systematic reviews met the inclusion criteria, providing data from 77 randomized controlled trials and 50 observational studies. All of them synthesized findings from low-quality primary studies, 2 of which used robust review methods. Most of the reviews focused on digital technologies to facilitate remote delivery of care, including consultations and therapy. No studies examined technologies used for first contact access to care, such as online appointment scheduling. Overall, we found no reviews of technology to facilitate first contact access to health and social care such as online appointment booking systems for older populations.

Conclusions: The impact of digital technologies on equitable access to services for older people is unclear. Research is urgently needed in order to understand the positive and negative consequences of digital technologies on health care access and to identify the groups most vulnerable to exclusion.

(J Med Internet Res 2021;23(11):e25887) doi:10.2196/25887

KEYWORDS

digital health; social care; access; older adults; review of reviews; umbrella review

Introduction

For at least a decade, the World Health Organization has encouraged member states to become leaders in serving citizens online, using digital technology to improve health and social care services [1]. Digital technologies are electronic tools, systems, and resources that generate, store, or process data [2]. The emergence of a novel coronavirus (SARS-CoV-2, which causes COVID-19) has led to the rapid rollout of digital technologies to support patient access to health and social care, while ensuring physical distancing [3]. Digital technologies were playing a growing role in connecting health and social care services with their users before the COVID-19 pandemic. A survey of patients aged 65 years and over in 9 countries in 2013 reported that over three-quarters preferred to book and manage their medical appointments online, and over three-quarters felt that online access to medical records was important [4]. The annual survey of 770,000 patients in UK family practice has described small increases in the proportion of people booking appointments (14.9% in 2019, up from 12.9% in 2018) and ordering repeat prescriptions online (16.2% in 2019, up from 14.3% in 2018) [5].

Supporting people to use digital health resources may help improve access to services, improve physical and mental well-being, and encourage shared decision-making [5]. However, estimates suggest that 37% of the world's estimated 7.8 billion population are digitally excluded [6], with older people, people on low incomes, and other marginalized groups most likely to be affected [5,7]. In the United States, around 80% of the population accesses the internet, but its use falls sharply with increasing age. Approximately 70% of the people aged 65 to 74 years are online, compared with 52% of those aged 75 to 84 years, and 38% aged ≥85 years [8]. In the United Kingdom, out of a total population of 66.4 million, approximately 11 million (20%) lack digital skills, and 8.4 million (8.5%) never go online [9], and just over half of the latter are aged over 65 [5]. There is a clear relationship between internet use and health, with increasing age, female gender, and greater deprivation being associated with lower internet use [10]. Potential barriers to digital access include lack of awareness, confidence, capacity, or skills [11,12], a reluctance to change established behaviors, and poor internet access [5]. Affordability and acceptability of digital technology is important in later life, and it is noteworthy that many devices have been developed without the involvement of older people [13]. The involvement of older adults in technological design and development can facilitate acceptability, although it is a complex matter and requires careful consideration [14].The recent widespread introduction of digital alternatives to face-to-face interactions makes it vital that we understand their impact on older adults' ability to access health and social care services that they need. In the United Kingdom, the National Health Service (NHS) roadmap sets out the milestones for digital health and social care to support people to live healthier lives and use fewer care services using technologies such as mobile phones and smartphones, tablets, and smart televisions [15]. It includes NHS digital health and wellbeing apps, such as the NHS app, which provides access to a range of NHS services via smartphones or tablets, and NHS login, which allows patients to view and access their personal health information online [16]. These technologies could potentially improve access to services by (1) facilitating first contact with services, (2) replacing face-to-face care with remote service delivery, and (3) providing access to professional support through remote patient monitoring [2]. Therefore, this review of reviews aims to answer the question of whether digital technologies improve access to health and social care for older adults and identify the characteristics of any digital interventions that are effective in increasing access to services for older adults.

Methods

Reporting Standards

We employed an umbrella review methodology to summarize the findings of previously published reviews [17]. The review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist for the reporting of systematic reviews [18]. The PRISMA checklist for this study is provided in Multimedia Appendix 1. Moreover, a review protocol was registered in the PROSPERO database [19].

Inclusion Criteria

The inclusion criteria were based on the PICOS (Population, Interventions, Comparator, Outcomes, and Study Designs) [20] criteria, which will be described in the following section.

Participants

Reviews of studies on older adults aged ≥65 or a combination of older and younger populations were selected in order to compare the effects of digital technologies on health care access between younger and older people.

Intervention

We used studies on any form of digital technology intended to facilitate access to appropriate health and social care services. These technologies enable first contact access (eg, online appointment scheduling) and are used as platforms for consultations and therapy interventions. They are also used in the remote care of patients. Furthermore, we recognized the fact that access to health and social care services would encompass availability and supply (ie, the degree of availability and quantity of supply at hand, regardless of whether they are used), utilization, equity, effectiveness, and quality of care [21].

Outcomes

We aimed to study the impact of digital technology on access to health and social care, which included the changes made in access and use of services as well as the cost-effectiveness of interventions that facilitate access and delivery of health and social care.

Study Designs

The study design of this paper encompassed any type of systematic review.

Search Strategy

We searched the following databases: Epistemonikos, MEDLINE (Ovid), Cochrane Database of Systematic Reviews (Wiley), ASSIA (ProQuest), PROSPERO, and for gray literature in Health Management Information Consortium (Ovid) and King's Fund. We used thesaurus headings along with title and abstract terms to search for digital technologies combined with specified outcomes for older people. The Canadian Agency for Drugs and Technologies in Health systematic review filter was adapted for databases that contained multiple study designs [22]. Searches were limited to the English language and the material published from January 1, 2000, to October 2019. The MEDLINE strategy is reported in Multimedia Appendix 2. The search results were downloaded to Endnote X9 (Clarivate Analytics) and deduplicated.

Data Collection

Two-stage screening was conducted by 2 reviewers independently using the Rayyan (Rayyan Systems) systematic review application [23]. We first tested and refined the inclusion and exclusion criteria on a sample of titles and abstracts to ensure that they were robust enough to capture relevant articles. The titles and abstracts of the reviews were screened against the refined inclusion criteria, followed by full text assessment of the selected articles. We resolved disagreements between the reviewers by discussion or by arbitration from another member of the review team.

Data Extraction

We extracted data into an Excel (Microsoft Corporation) spreadsheet, using a form based on the Cochrane Data Extraction and Assessment Template [20] to record the relevant review characteristics. The extracted data included: (1) author and year of publication; (2) title; (3) objective of the review; (4) description of the included population; (5) total number of older people; (6) intervention; (7) technology type; (8) what the intervention is enhancing; (9) primary outcomes; (10) secondary outcomes; (11) overall statement on quality appraisal; and (12) review authors' summary. To ensure comprehensiveness, we piloted the abstraction form on 2 reviews, which identified a need for minor modifications. Risk of bias was assessed using the ROBIS (Risk of Bias in Systematic Reviews) tool [24]. We chose to use ROBIS as opposed to AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) because we are experienced with the former, and a comparative analysis of the two tools showed little difference between them [25].

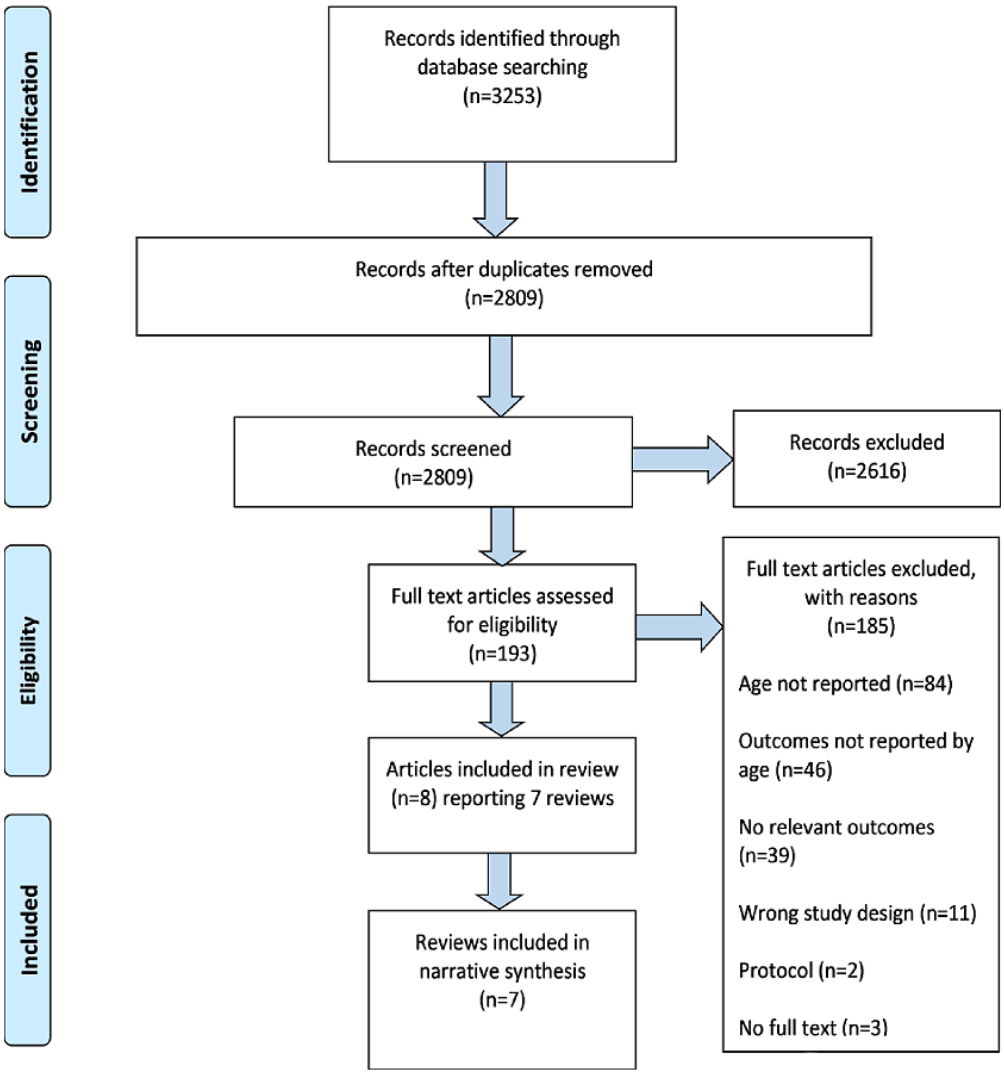
Data Analysis

We presented our main results in tabular format with a narrative synthesis. We grouped the results according to the three purposes of digital health technology, which consist of enabling first contact access, consultations and therapy, and remote monitoring. Due to a lack of data, we were unable to analyze the effects of interventions at ages over 65 years.

Results

Database searches identified 2809 unique records. The initial screening of title and abstracts excluded 2616 records, leaving 193 for full text assessment (Figure 1). We identified 7 reviews eligible for inclusion.

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



Characteristics of the Included Reviews

A total of 7 reviews published between 2006 and 2019 met the inclusion criteria [26-32]. A descriptive summary of review characteristics is presented in Table 1. The 7 included reviews include a total of 77 randomized controlled trials (RCTs) and 50 observational studies. We assessed the overlap across the reviews and identified 7 RCTs [33-39] reported in more than 1 review, but no observational studies that were included in more than 1 review. The studies in the reviews included 49 from the United States, 40 from Europe (including 7 from the United Kingdom), 9 from Australia, 6 from Canada, and 6 from the rest of the world. Country of origin was not stated for the remaining 17 studies. All of the studies reported outcomes for adults aged 65 and older, and 2 reviews included adults from age 18 [26,32].

Table 1. Summary of included systematic reviews.

Author	Study designs included in the review	Population	Intervention	Type of technology	Outcome
Bauce [26]	RCT,[a] observational	Adults aged >65 (in 10 out of 11 studies)	Telemonitoring	Videophones, smartphone, and mobile phone	Hospital admissions and emergency department visits
Harerimana [27]	RCT, observational	Adults aged ≥65 with a diagnosis of depression or self-reported depressive symptoms	Telehealth (mental health)	Telephone and computers	Hospital admissions and emergency department visits
Husebo [28]	Observational	Adults aged >65, either living alone or receiving informal care	Telehealth	Videophones, personal computers or laptops, and TV	Hospital admissions and readmissions
Inglis [29,40]	RCT	Adults with heart failure (8 studies included people with a mean age of ≥70)	Structured telephone support or telemonitoring (heart failure)	Telephone	Heart failure and all-cause hospitalizations
Martinez [30]	RCT, observational	Adults with heart failure (11 studies included people with a mean age of ≥65)	Home telecare	Not reported	Hospital readmissions
Marx [31]	RCT, observational	Adults with a mean age of ≥65 years living independently, in receipt of intervention for management risk of malnutrition	Telehealth for managing risk of malnutrition	Telephone and computer	Hospital readmission and healthcare costs
Sanyal [32]	RCT, observational	Older adults (11 studies included people with a mean age of ≥65 years)	Telehealth, cognitive behavior therapy	Computer	Cost-effectiveness or utility of eHealth technologies

[a]RCT: randomized controlled trial.

Risk of Bias Assessment

Details of the risk of bias assessment can be found in Table 2. Overall, the risk of bias was high for 5 reviews [26-28,30,32], and low for 2 reviews [29,31]. The main issues were the absence of clear inclusion criteria and the lack of publicly available protocols with predefined criteria. A detailed description of risk of bias assessment is reported in Multimedia Appendix 3.

Table 2. Risk of bias using ROBIS (Risk of Bias in Systematic Reviews) assessment.

Review	Phase 2				Phase 3
	Study Eligibility Criteria	Identification and selection of studies	Data collection and study appraisal	Synthesis and findings	Overall risk of bias
Bauce [26]	High	Unclear	High	High	High
Harerimana [27]	High	Unclear	High	High	High
Husebo [28]	High	High	High	High	High
Inglis [29,40]	Low	Low	Low	Low	Low
Martinez [30]	High	High	High	High	High
Marx [31]	Low	Low	Low	Low	Low
Sanyal [32]	High	Low	Unclear	High	High

Outcomes

Table 3 summarizes the identified evidence, presenting it according to the purpose of the digital technology and the reported outcomes. None of the reviews reported outcomes that were related to the changes in access to services. In total, 6 reviews reported on hospital admissions [26-31], 1 reported on healthcare costs [31], and 1 on the cost-effectiveness of digital technology [32]. A variety of digital technologies were used by healthcare professionals and older adults to support interventions for telemonitoring or telecare: videophones or video conferencing equipment, internet-based applications, and smartphones.

Table 3. Overview of the identified evidence by type of digital technology and outcome.

Objective	Outcome	
Purpose of digital technology	Health service utilization	Costs and cost-effectiveness
Digital technology to enable first point of contact access (eg, online appointment scheduling)	No reviews identified	No reviews identified
Digital technologies or platforms for consultations and therapy interventions	Harerimana [27]; Marx [31]; Husebo [28]; Inglis [29]; Martinez [30]	Sanyal [32]
Digital technology for remote monitoring interventions	Bauce [26]	Sanyal [32]

First Point of Contact Access

No systematic reviews reported evidence about the impact of digital technology to facilitate first point of contact access with health services, such as online appointment scheduling.

Consultations and Therapies

In total, 5 reviews reported on health care service utilization, in malnutrition [31], heart failure [29,30], and mental health [27,28], as outcomes of digital technologies, but only 2 reviews were judged to be at low risk of bias and thus of higher quality [29,31].

Malnutrition

Marx and colleagues [31] reported weak evidence for the effectiveness of telehealth interventions to address malnutrition among community-dwelling older adults. They identified 9 studies (7 RCTs and 2 observational); 2 of the 9 studies reported significant reductions in hospital readmissions in the intervention groups. However, when the data were pooled, the reduction in hospital admissions was not significant; (odds ratio 0.52, 95% CI 0.24-1.16); *P*=.11; n=160; I^2=0%).

Heart Failure

Inglis and colleagues [29] focused on whether structured telephone support and telemonitoring were effective for older people with heart failure. They found 41 RCTs that assessed heart failure–related hospitalizations. A meta-analysis of some of the included studies reported a 15% reduction in risk for heart failure–related hospitalizations with structured telephone support (relative risk 0.85, 95% CI 0.77-0.93; n=7030; 16 studies; I^2=27%) and a 29% reduction in telemonitoring (relative risk 0.71, 95% CI 0.60 to 0.83; n=2148; 8 studies; I^2=20%). There were no impacts reported on all-cause hospitalizations. The quality of the evidence reported for these heart failures and all-cause hospitalization studies was rated very low [29]. Evidence from the lower-quality reviews reported positive impacts of digital technology interventions on service utilization. Martinez and colleagues [30] reviewed 42 articles on the value of home monitoring for heart failure patients, 5 of which reported findings for older people. Remote consultations and follow-up care were associated with lower admission and readmission rates.

Mental Health

Husebo and colleagues [28] sought to understand the care content and utilization of virtual visits, particularly the uses and experiences of adults aged 65 and over. In their review, 1 study reported that all-cause readmissions were lower in the telehealth group (n=102) compared with standard care (n=116). At 30 days, 16 (16%) versus 22 (19%) and at over 6 months, 46 (46%) versus 60 (52%) of intervention versus control patients were readmitted [41]. Telehealth has also been used to deliver mental health care for older adults with depressive symptoms (telemental health). In a 6-month single (quasi-experimental) study of 76 patients, identified by Harerimana and colleagues [27], telemental health reduced hospital admissions by 80% (46 versus 9 admissions) and emergency room visits 60% (80 versus 32 visits) [42]. Evidence for the impact of digital technologies on economic outcomes was sparse. Moreover, 1 single review of eHealth technologies in the management of chronic diseases reported limited evidence, which did not support the assessment of cost-effectiveness [32].

Remote Monitoring

Two reviews reported evidence about technologies for remote monitoring, both of which were judged to be of poor quality. Bauce and colleagues [26] assessed the effectiveness of telemonitoring (videoconferencing) interventions on heart failure outcomes in 11 studies (10 RCTs and 1 single-group study). Five studies reported significant reductions in hospital admissions, and 2 others reported significant reductions in emergency department visits. The authors speculated that the reduction in healthcare use was likely to be due to the early detection and treatment of symptoms attributable to the intervention. Reduction in hospital admissions due to telemonitoring was supported by Queirós and colleagues [43]. Their systematic review assessed the use of technologies in the remote care of patients with long-term conditions such as diabetes, congestive heart failure, chronic obstructive pulmonary disease, and mental disorders [43].

Discussion

Principal Results

We identified evidence on a variety of digital technologies to facilitate interaction between older people and services at different parts of the care pathway. However, we found no reviews of technology to facilitate first point of contact access such as online appointment booking systems. There was no significant difference in hospital admissions for telehealth interventions (but this may have been due to the studies' lack of power as there were only 160 participants in the pooled analysis) [31]. However, for heart failure, structured telephone support resulted in 15% reduction in admissions [29]. Other reviews were of too low a quality to permit confidence in

findings, however, and there were no signs that a focus on reviews with too low a risk of bias would change anything. From the 7 overlapping RCTs [33-39], benefits to the older population in access were poorly measured and not clearly reported. In these RCTs, focus was on reducing hospital admissions, and there was little account of whether these technologies are enabling older people to interact with or access health and social care services more effectively. There was also no review evidence for newer technologies such as smartphone apps (eg, the NHS app in the United Kingdom), some of which were already in widespread use before COVID-19 [15].

The 2020 COVID-19 pandemic prompted the rapid implementation of alternatives to face-to-face interactions in health and social care [3]. This was a pragmatic response to a novel emergency that allowed care delivery to continue. As the pandemic evolves, digital innovations that have been implemented at speed should be evaluated to ensure that they are effective and affordable so that they can promote equitable access and do not selectively overlook certain sections of the population [14]. However, none of the included reviews addressed the issue of affordability and acceptability of digital technology in later life. For sections of the population who lack digital literacy or a means of digital engagement, the benefits are less clear, and there is every possibility that they will be harmed by losing the ability to access services in traditional, nondigital ways.

Most of the evidence [26-31] was concerned with digital technologies to facilitate remote delivery of care, including consultations and therapy, reflecting a research focus congruent with policy priorities [15]. However, these evaluations were more focused on reducing hospital utilization rather than enhancing access to services. Whether digital technologies do reduce hospital admissions and visits by facilitating timely access to appropriate alternative care is impossible to determine from the evidence presented. Evidence on the cost-effectiveness of digital health technologies was confined to 1 low-quality review, from which no clear conclusions can be drawn [32].

Limitations

To the best of our knowledge, this is the first rapid synthesis of systematic reviews on digital technology aimed at enhancing access to health and social care services for older adults. We followed a rapid evidence synthesis approach, and our database searching, handling of data, and reporting adhered to published guidelines for undertaking a robust standard systematic review [18,44]. We restricted our searches to English language publications due to time constraints and acknowledge that this may have excluded relevant material. Two limitations of the material should be highlighted. First, most of the studies contained within our included reviews were randomized trials of effectiveness and cost-effectiveness. However, we found that the benefits to the older population in access are poorly measured and not clearly reported in studies of digital technology. Second, most of the reviews failed to adequately report their findings, and formal assessments of the methodological quality indicated a low-quality evidence base. This leads us to be cautious in our interpretation of the evidence and any conclusions drawn.

Comparison With Prior Work

Our assessment of the dearth of evidence on first point of contact digital technology is supported by other works. A recent review of approaches to the evaluation of digital health interventions identified little evidence from randomized controlled trials and carried out measurement of service utilization in only a minority of the studies [45]. Our review suggests that digital health technologies may be associated with reductions in health service use. This is supported by multiple systematic reviews in younger populations of patients with long-term conditions [43]. There is a particular gap in the evaluation of any digital technologies used in social care.

Implications for Policy, Research, and Practice

The COVID-19 pandemic has resulted in the rapid implementation of digital interventions to allow continued access to services when infection risk was high. This rapid rollout went beyond any evidence for effectiveness, driven by the extraordinary need to reduce face-to-face contact. However, prepandemic concerns about the adverse effects of digital technologies on access to services for older people remain valid. For older people who are digitally excluded, these digital interventions risk exacerbating any problems they already faced when trying to access health and social care services. This, in turn, has implications for workload in primary care, and health care providers must take on greater responsibility to ensure that this important section of the population receives the care it needs. There is a notable gap in the evidence for studies assessing the impact of technologies to enable first point of contact for health and social care services (eg, online platforms to book appointments). A mapping review of primary studies is required to understand this impact on different population subgroups, but this is unlikely to be sufficient. Further work is needed to understand the effectiveness and cost-effectiveness of digital technologies and their effect on equity of access to health and social care services. This should encompass access to appropriate care, which may lead to reductions in the use of other services as well as changes in health outcomes. The paucity of evidence in this area points to the need for a broad research program in partnership with older people and service providers in order to understand the characteristics of digital technologies, which can enhance access to services.

Conclusions

The current systematic review evidence on the potential for digital technologies to improve access to health and social care for older adults is limited in both scope and quality. However, these limited attempts raise the possibility that providing digital interventions in addition to or as a replacement for face-to-face services may reduce demands on hospitals. Further research is required, and the widespread use of digital technologies to facilitate access to health and social care during the COVID-19 pandemic offers an ideal opportunity to better understand the barriers, facilitators, and limitations of their use.

Acknowledgments

We thank other members of the National Institute for Health Research (NIHR) Policy Research Unit in Older People and Frailty with whom we discussed the ideas presented in this paper during unit meetings. This paper presents independent research funded by the NIHR Policy Research Unit in Older People and Frailty. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[PDF File (Adobe PDF File), 395 KB - jmir_v23i11e25887_app1.pdf]

Multimedia Appendix 2
Example of search strategy.
[DOCX File , 18 KB - jmir_v23i11e25887_app2.docx]

Multimedia Appendix 3
Risk of bias assessment of the included reviews.
[DOCX File , 17 KB - jmir_v23i11e25887_app3.docx]

References

1. Draft global strategy on digital health 2020-2024. World Health Organization. 2019. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_42 [accessed 2020-05-18]
2. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 Dec 07;37(18):1428-1438 [FREE Full text] [doi: 10.1093/eurheartj/ehv770] [Medline: 26873093]
3. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. The Lancet Digital Health 2020 Aug;2(8):e425-e434 [FREE Full text] [doi: 10.1016/S2589-7500(20)30137-0]
4. 2013 Accenture Consumer Survey on Patient Engagement. Accenture. 2013. URL: https://www.criticaleye.com/inspiring/insights-servfile.cfm?id=3869 [accessed 2020-11-15]
5. Digital inclusion for health and social care: What we mean by digital inclusion. National Health Service. 2019. URL: https://digital.nhs.uk/about-nhs-digital/our-work/digital-inclusion/what-digital-inclusion-is [accessed 2020-01-27]
6. World Internet Users and 2021 Population Stats. Internet World Stats. 2020. URL: https://www.internetworldstats.com/stats.htm [accessed 2020-05-20]
7. Leist AK. Social media use of older adults: a mini-review. Gerontology 2013;59(4):378-384 [FREE Full text] [doi: 10.1159/000346818] [Medline: 23594915]
8. NTIA Data Reveal Shifts in Technology Use, Persistent Digital Divide. National Telecommunications and Information Administration. 2020. URL: https://www.ntia.doc.gov/blog/2020/ntia-data-reveal-shifts-technology-use-persistent-digital-divide [accessed 2020-07-21]
9. Estimates of the population for the UK. Office for National Statistics. 2017. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland [accessed 2020-01-27]
10. Matthews K, Nazroo J, Marshall A. Digital inclusion in later life: cohort changes in internet use over a ten-year period in England. Ageing and Society 2018 Aug 02;39(9):1914-1932. [doi: 10.1017/s0144686x18000326]
11. Measuring the Information Society Report 2018. International Telecommunications Union. 2018. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/misr2018.aspx [accessed 2020-05-18]
12. GP Patient Survey 2019. National Health Service. 2019. URL: https://www.england.nhs.uk/statistics/2019/07/11/gp-patient-survey-2019/ [accessed 2020-01-27]
13. Connected Health: How digital technology is transforming health and social care. The Deloitte Centre for Health Soultions. 2015. URL: https://www2.deloitte.com/content/dam/Deloitte/ie/Documents/PublicSector/deloitte-uk-connected-health-sm1.pdf [accessed 2019-12-18]
14. 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]

15. The Forward View into Action: Paper-free at the Point of Care Guidance for Developing Local Digital Roadmaps. National Health Service. 2016. URL: https://www.england.nhs.uk/digitaltechnology-old/wp-content/uploads/sites/31/2016/11/develp-ldrs-guid.pdf [accessed 2020-01-27]
16. Apps and tools for patient care. NHSX. 2020 Jan 29. URL: https://www.nhsx.nhs.uk/key-information-and-tools/apps-and-tools [accessed 2020-01-29]
17. Papatheodorou S. Umbrella reviews: what they are and why we need them. Eur J Epidemiol 2019 Jun;34(6):543-546. [doi: 10.1007/s10654-019-00505-6] [Medline: 30852716]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
19. Kunonga P, Spiers G, Beyer F, Boulton E, Hall A, Craig D. Digital technologies providing access to health and social care: a rapid evidence synthesis. PROSPERO: International prospective register of systematic reviews. 2019. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=161421 [accessed 2020-01-12]
20. CRD's Guidance for Undertaking Reviews in Health Care. Centre for Reviews and Dissemination. 2009. URL: https://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm [accessed 2019-11-20]
21. Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, et al. What does 'access to health care' mean? J Health Serv Res Policy 2002 Jul;7(3):186-188. [doi: 10.1258/135581902760082517] [Medline: 12171751]
22. Strings Attached: CADTH's Database Search Filters. Canadian Agency for Drugs and Technologies in Health. 2010. URL: https://www.cadth.ca/resources/finding-evidence/strings-attached-cadths-database-search-filters#syst [accessed 2020-01-01]
23. 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]
24. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, ROBIS group. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016 Jan;69:225-234 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.06.005] [Medline: 26092286]
25. Pieper D, Puljak L, González-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. J Clin Epidemiol 2019 Apr;108:26-33. [doi: 10.1016/j.jclinepi.2018.12.004] [Medline: 30543911]
26. Bauce K, Fahs DB, Batten J, Whittemore R. Videoconferencing for Management of Heart Failure: An Integrative Review. J Gerontol Nurs 2018 Apr 01;44(4):45-52. [doi: 10.3928/00989134-20180207-01] [Medline: 29437182]
27. Harerimana B, Forchuk C, O'Regan T. The use of technology for mental healthcare delivery among older adults with depressive symptoms: A systematic literature review. Int J Ment Health Nurs 2019 Jun;28(3):657-670. [doi: 10.1111/inm.12571] [Medline: 30666762]
28. Husebø AML, Storm M. Virtual visits in home health care for older adults. ScientificWorldJournal 2014;2014:689873 [FREE Full text] [doi: 10.1155/2014/689873] [Medline: 25506616]
29. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
30. Martínez A, Everss E, Rojo-Alvarez JL, Figal DP, García-Alberola A. A systematic review of the literature on home monitoring for patients with heart failure. J Telemed Telecare 2006;12(5):234-241. [doi: 10.1258/135763306777889109] [Medline: 16848935]
31. Marx W, Kelly JT, Crichton M, Craven D, Collins J, Mackay H, et al. Is telehealth effective in managing malnutrition in community-dwelling older adults? A systematic review and meta-analysis. Maturitas 2018 May;111:31-46. [doi: 10.1016/j.maturitas.2018.02.012] [Medline: 29673830]
32. 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]
33. Cleland JGF, Louis AA, Rigby AS, Janssens U, Balk AHMM, TEN-HMS Investigators. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005 May 17;45(10):1654-1664 [FREE Full text] [doi: 10.1016/j.jacc.2005.01.050] [Medline: 15893183]
34. de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring. The report of a pilot study of patients with chronic heart failure. European Journal of Heart Failure 2001 Dec 03;3(6):723-730 [FREE Full text] [doi: 10.1016/s1388-9842(01)00190-8] [Medline: 11738225]
35. Dimmick SL, Burgiss SG, Robbins S, Black D, Jarnagin B, Anders M. Outcomes of an integrated telehealth network demonstration project. Telemed J E Health 2003;9(1):13-23. [doi: 10.1089/153056203763317611] [Medline: 12699604]
36. Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL, et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: the Weight Monitoring in Heart Failure (WHARF) trial. American Heart Journal 2003 Oct;146(4):705-712. [doi: 10.1016/S0002-8703(03)00393-4] [Medline: 14564327]

37. Jerant AF, Azari R, Nesbitt TS. Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention. Med Care 2001 Nov;39(11):1234-1245. [doi: 10.1097/00005650-200111000-00010] [Medline: 11606877]
38. Wakefield BJ, Ward MM, Holman JE, Ray A, Scherubel M, Burns TL, et al. Evaluation of home telehealth following hospitalization for heart failure: a randomized trial. Telemed J E Health 2008 Oct;14(8):753-761. [doi: 10.1089/tmj.2007.0131] [Medline: 18954244]
39. Woodend A, Fraser M, Sprang M, Stueve L. Readmission and quality of life: the impact of telehomecare in heart failure. J Card Fail 2002;8:98 [FREE Full text]
40. Inglis SC, Conway A, Cleland JG, Clark RA. Is age a factor in the success or failure of remote monitoring in heart failure? Telemonitoring and structured telephone support in elderly heart failure patients. Eur J Cardiovasc Nurs 2015 Jun;14(3):248-255. [doi: 10.1177/1474515114530611] [Medline: 24681423]
41. Bowles KH, Hanlon AL, Glick HA, Naylor MD, O'Connor M, Riegel B, et al. Clinical effectiveness, access to, and satisfaction with care using a telehomecare substitution intervention: a randomized controlled trial. Int J Telemed Appl 2011;2011:540138 [FREE Full text] [doi: 10.1155/2011/540138] [Medline: 22187551]
42. Godleski L, Cervone D, Vogel D, Rooney M. Home telemental health implementation and outcomes using electronic messaging. J Telemed Telecare 2012 Jan;18(1):17-19. [doi: 10.1258/jtt.2011.100919] [Medline: 22052966]
43. Queirós A, Alvarelhão J, Cerqueira M, Silva A, Santos M, Pacheco Rocha N. Remote Care Technology: A Systematic Review of Reviews and Meta-Analyses. Technologies 2018 Feb 10;6(1):22. [doi: 10.3390/technologies6010022]
44. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. Syst Rev 2018 Mar 01;7(1):39 [FREE Full text] [doi: 10.1186/s13643-018-0695-8] [Medline: 29490699]
45. Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon M. Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews. J Med Internet Res 2018 Jun 07;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
NHS: National Health Service
NIHR: National Institute for Health Research
PICOS: Population, Interventions, Comparator, Outcomes, and Study Designs
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized control trial
ROBIS: Risk of Bias in Systematic Reviews

Edited by R Kukafka, G Eysenbach; submitted 19.11.20; peer-reviewed by M O'Connell, B Davis, H Mehdizadeh, J Delgado-Ron; comments to author 19.01.21; revised version received 23.02.21; accepted 15.10.21; published 24.11.21.

Please cite as:
Kunonga TP, Spiers GF, Beyer FR, Hanratty B, Boulton E, Hall A, Bower P, Todd C, Craig D
Effects of Digital Technologies on Older People's Access to Health and Social Care: Umbrella Review
J Med Internet Res 2021;23(11):e25887
URL: https://www.jmir.org/2021/11/e25887
doi:10.2196/25887
PMID:34821564

Original Paper

Prediction of Smoking Risk From Repeated Sampling of Environmental Images: Model Validation

Matthew M Engelhard[1], MD, PhD; Joshua D'Arcy[1], MEng, MD; Jason A Oliver[1], PhD; Rachel Kozink[1], MSc; F Joseph McClernon[1], PhD

Department of Biostatistics & Bioinformatics, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Matthew M Engelhard, MD, PhD
Department of Biostatistics & Bioinformatics
Duke University School of Medicine
2608 Erwin Rd
Durham, NC, 27705
United States
Phone: 1 919 613 3665
Email: m.engelhard@duke.edu

Abstract

Background: Viewing their habitual smoking environments increases smokers' craving and smoking behaviors in laboratory settings. A deep learning approach can differentiate between habitual smoking versus nonsmoking environments, suggesting that it may be possible to predict environment-associated smoking risk from continuously acquired images of smokers' daily environments.

Objective: In this study, we aim to predict environment-associated risk from continuously acquired images of smokers' daily environments. We also aim to understand how model performance varies by location type, as reported by participants.

Methods: Smokers from Durham, North Carolina and surrounding areas completed ecological momentary assessments both immediately after smoking and at randomly selected times throughout the day for 2 weeks. At each assessment, participants took a picture of their current environment and completed a questionnaire on smoking, craving, and the environmental setting. A convolutional neural network–based model was trained to predict smoking, craving, whether smoking was permitted in the current environment and whether the participant was outside based on images of participants' daily environments, the time since their last cigarette, and baseline data on daily smoking habits. Prediction performance, quantified using the area under the receiver operating characteristic curve (AUC) and average precision (AP), was assessed for out-of-sample prediction as well as personalized models trained on images from days 1 to 10. The models were optimized for mobile devices and implemented as a smartphone app.

Results: A total of 48 participants completed the study, and 8008 images were acquired. The personalized models were highly effective in predicting smoking risk (AUC=0.827; AP=0.882), craving (AUC=0.837; AP=0.798), whether smoking was permitted in the current environment (AUC=0.932; AP=0.981), and whether the participant was outside (AUC=0.977; AP=0.956). The out-of-sample models were also effective in predicting smoking risk (AUC=0.723; AP=0.785), whether smoking was permitted in the current environment (AUC=0.815; AP=0.937), and whether the participant was outside (AUC=0.949; AP=0.922); however, they were not effective in predicting craving (AUC=0.522; AP=0.427). Omitting image features reduced AUC by over 0.1 when predicting all outcomes except craving. Prediction of smoking was more effective for participants whose self-reported location type was more variable (Spearman ρ=0.48; *P*=.001).

Conclusions: Images of daily environments can be used to effectively predict smoking risk. Model personalization, achieved by incorporating information about daily smoking habits and training on participant-specific images, further improves prediction performance. Environment-associated smoking risk can be assessed in real time on a mobile device and can be incorporated into device-based smoking cessation interventions.

(J Med Internet Res 2021;23(11):e27875) doi:10.2196/27875

KEYWORDS

smoking; smoking cessation; machine learning; computer vision; digital health; eHealth; behavior; CNN; neural network; artificial intelligence; AI; images; environment; ecological momentary assessment; mobile health; mHealth; mobile phone

Introduction

Background

Cigarette smoking is the leading cause of preventable deaths in the United States [1], and tobacco use is responsible for more than 7 million annual deaths worldwide [2]. Although most smokers are motivated to quit [3], fewer than 10% of quit attempts are successful [4], which has motivated ongoing efforts to develop more effective cessation strategies.

New, mobile device-based cessation interventions have improved 6-month [5] and 12-month [6] cessation outcomes, and success rates can be further improved [7] by tailoring interventions to individual users and momentary contexts, such as the user's geographic location [8]. This paradigm has been formalized as the just-in-time adaptive intervention (JITAI) [9,10], wherein a cessation support system continuously monitors contextual factors through ecological momentary assessment (EMA; ie, repeated self-reporting), passive sensing, or a combination of the two, then provides context-sensitive support to smokers at times when it is most needed. By monitoring user physiology [11-14], geographic location [8,15], and recent smoking events [15], a cessation support system can estimate smoking risk from moment to moment and then intervene when the estimated risk is high.

However, most smoking cessation interventions neglect an important contributor to smoking behaviors—the smoker's external environment. A growing body of evidence, collected in both laboratory and real-world settings, suggests that smoking risk is affected not only by internal factors but also by the smoker's current environmental context. For example, images of personal smoking environments increase smoking behaviors and self-reported craving [16-19] and activate neural circuits associated with craving, contributing to subsequent smoking behaviors [20]. Consistent with these findings, a recent study showed that self-reported environmental conditions, such as being around other smokers or in a place where smoking is permitted, were stronger predictors of smoking lapse than internal states (eg, smoking urge) [21].

Current technologies can now quantify the effects of environmental factors on real-world smoking behaviors [22]. Wearable cameras and mobile devices allow images of daily environments to be acquired near continuously at low cost [23-26], and computer vision models, now highly accurate and optimized for mobile devices [27], can identify environmental features in these images. These features can then be linked to smoking and craving, along with other behaviors or outcomes of interest. This process is objective, does not require manual annotation of images, and scales to large data sets and study populations.

In a previous study, we demonstrated that computer vision could distinguish between daily environments where smokers commonly smoke and those where they rarely smoke. Using the approach outlined above, we also uncovered specific objects and settings associated with smoking versus nonsmoking environments [28]. These findings suggest that environmental features monitored via computer vision may provide important contextual information that can improve the prediction of momentary smoking risk. However, these two extremes, known smoking and nonsmoking environments, do not reflect the full range of environments that smokers encounter in their daily lives.

Objective

In this study, we collected a representative sample of images of smokers' daily environments through photograph-augmented EMA (photoEMA). In each assessment, participants self-reported recent smoking and their current craving level and then took a picture of their environment. A mobile-optimized convolutional neural network was trained to predict smoking risk and other outcomes relevant to smoking (craving, whether smoking was permitted in the current environment, and whether the participant was outside) based on environmental images and other participant-specific features. We hypothesized that out-of-sample prediction would be effective, providing a basis for an environment-aware JITAI, and that prediction performance could be improved through model personalization, in which images from a given participant are used to refine model predictions for that participant. We also aim to understand how model performance varies by location type, as reported by participants. Our final prediction model, QuitEye, was deployed on a mobile device and can assess environment-associated smoking risk and craving in real time to support environment-aware smoking cessation interventions.

Methods

Study and Participants

Recruitment and all study procedures were approved by the Duke University Health System Institutional Review Board, and written consent was obtained from all participants. Smokers (≥10 cigarettes per day for ≥2 years) aged ≥18 years were recruited from the Durham, North Carolina area. Participants were recruited from the community for a study of smoking behavior via printed and web advertisements and word-of-mouth. Participants were excluded if they regularly used noncigarette tobacco products (eg, e-cigarettes); currently used smoking cessation medications; planned to quit smoking, otherwise altered their smoking pattern, left the study area or anticipated a major life event during the study; had current or recent alcohol or drug abuse problems; or were pregnant, breastfeeding, or planning to become pregnant during the study. Eligible participants completed an initial visit to (1) biochemically verify their smoking status (ie, carbon monoxide breath test) and test for illicit drug use, (2) test for pregnancy, and (3) complete questionnaires on nicotine dependence and tobacco use history. Participants who met all eligibility requirements (n=52) then downloaded the photoEMA app

(Metricwire) to their smartphone and were trained on its use. Following the 14-day photoEMA period, participants completed a follow-up visit during which an interview was conducted to assess drug and alcohol use, tobacco purchasing, and any other events that might have affected smoking (eg, illness) or daily living (eg, death in the family) patterns. Participants were compensated for up to US $350 in total, including daily (US $5) and weekly (US $50) incentives for high photoEMA completion. All procedures were observational, and no randomization or intervention was performed.

PhotoEMA Collection

Participants completed the photoEMA assessments for 14 days.

Random Prompts

Participants specified their typical wakeful hours during the screening. They were prompted six times daily at randomly spaced intervals. The average interval between prompts was 120 minutes in duration. At each assessment, participants rated their current levels of urge to smoke (1 item) and affect and stress (11 items; not reported here). In addition, they captured a time-stamped image of their current location. Finally, they were prompted to label the location with a prepopulated list of common locations (eg, bedroom, office, car, or park), other location information (eg, indoors or outdoors and whether smoking was permitted), current activity (eg, working or running errands), social environment (eg, presence of others), and recent alcohol and caffeine use.

Smoking-Initiated Assessments

Participants were also instructed to complete assessments each time they smoked. They were asked how many cigarettes they smoked in this location on this occasion and all items from the random prompt assessments.

Across Prompt Types

Participants were instructed to delay responding if they were in situations or locations where responding to, or initiating prompts, would be distracting (eg, in a meeting) or dangerous (eg, while driving). Across assessments, participants were asked to compose pictures to avoid including other people but to otherwise leave environments as they are.

Craving data were dichotomized based on the median self-reported craving for all participants. Self-reported craving of a *moderate* or lower level was coded as negative, and self-reported craving of *quite a bit* or higher level was coded as positive. Other outcomes (smoking, whether smoking was permitted, and whether the participant was outside) were binary; therefore, no binarization was required.

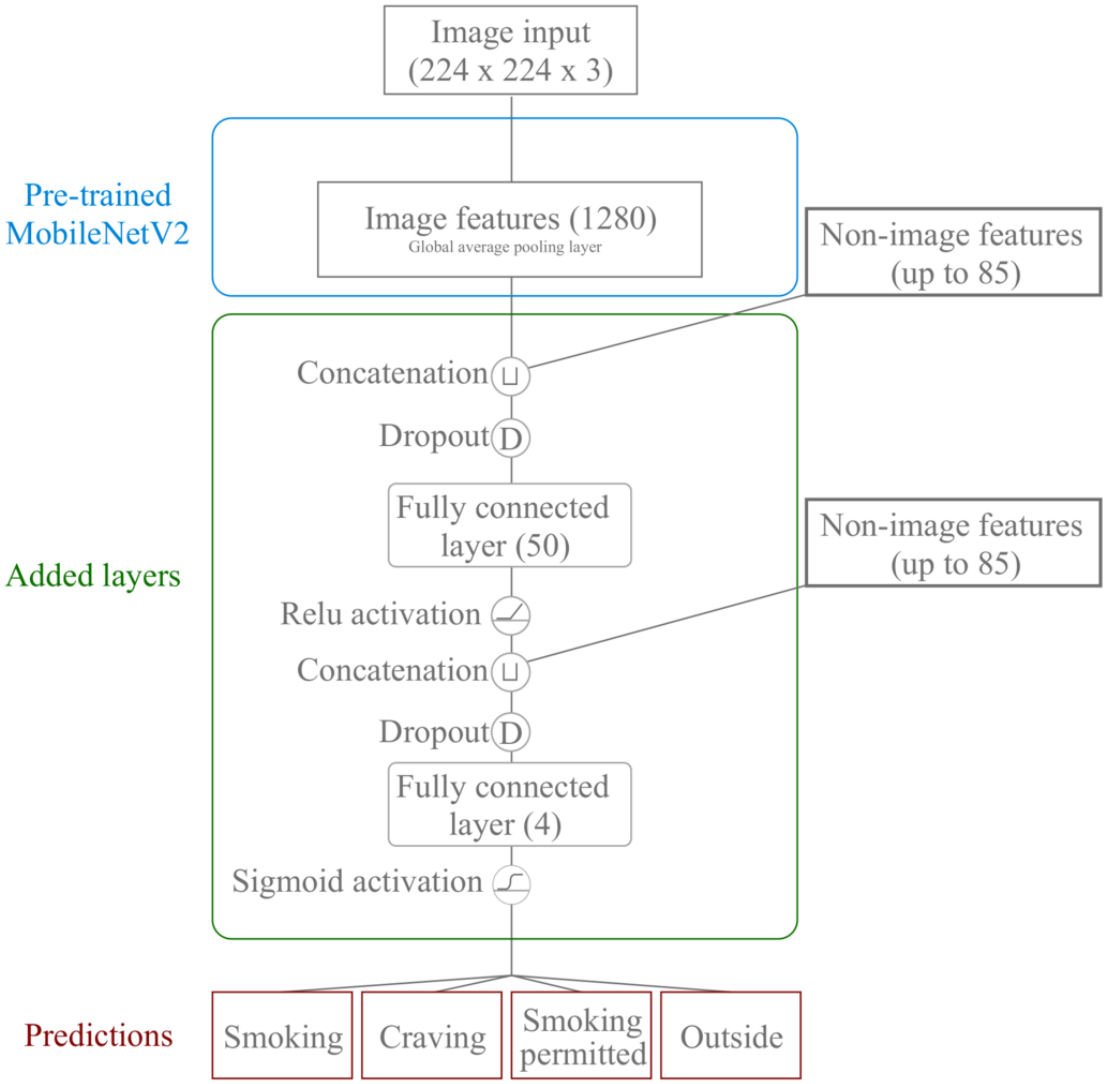
Convolutional Neural Network Model

QuitEye is based on MobileNetV2, a convolutional neural network architecture optimized for mobile devices [27] pretrained on ImageNet [29]. QuitEye also incorporates the following additional information: (1) participant age and sex, (2) known smoking locations inside and outside the home as indicated by self-report at baseline, (3) time since the last cigarette, (4) time of day and day of the week, and (5) a participant-specific indicator variable (in the longitudinal models only).

To determine the impact of each of these elements on prediction performance, we conducted ablation studies in which models *that did not incorporate* a given element were also trained and evaluated. For example, we trained a model *without* age and sex to assess the impact of these factors on prediction performance, and to estimate performance for a possible deployment in which these were not available. Models that did not incorporate image features were also trained and evaluated to determine the impact of the images on prediction performance.

QuitEye is a multi-task architecture that jointly predicts four binary outcomes: smoking, craving, whether smoking is permitted, and whether the participant is outside. Prediction of whether the participant is outside was included both to contextualize other performance figures and because inside or outside status is associated with smoking behaviors. Nonimage features were concatenated with image features from the global pool layer of MobileNetV2, and a single hidden neural network layer (rectified linear unit activation) was applied. Nonimage features were again concatenated to the output of this layer, and a second fully connected layer (sigmoid activation) was then used to predict each of the four binary outcomes. The QuitEye architecture is shown in Figure 1.

Figure 1. Diagram of QuitEye, which extracts image features using the MobileNetV2 convolutional neural network, then predicts smoking status, craving, whether smoking is permitted, and whether the participant is outside based on a combination of image features and additional data collected from participants with a mobile device.



Training and Evaluation

QuitEye was trained using Tensorflow v1.15 in Python v3.7 on a single Titan XP GPU. MobileNetV2 parameters were initialized to values learned on ImageNet [29], and all parameters were fine-tuned. Two training and evaluation procedures were used to evaluate out-of-sample performance (ie, *out-of-the-box* performance, without first learning from images from a given participant) and personalized model performance (ie, after learning from a subset of that participant's images).

Out-of-sample performance was assessed by training and evaluating the model using nested cross-validation [30] with five outer folds and five inner folds. In the nested cross-validation procedure, out-of-sample performance was evaluated on each outer fold after developing the model using data from the remaining folds. In each development set, an inner cross-validation procedure was used to determine the optimal hyperparameter settings.

Personalized model performance was assessed by developing the model with data from all participants from days 1 to 10, then evaluating it on data from days 11 to 14. Images used in model development were divided at random into training (80%) and validation (20%) sets.

Hyperparameters included the width of the hidden layer (Figure 1), hidden layer dropout rate, and learning rate. Performance was evaluated using the area under the receiver operating characteristic curve (AUC) for each of the four prediction tasks.

Additional models were trained using the procedures outlined above to quantify the impact of additional (nonimage) features on performance. Features were categorized as (1) baseline information, including participant demographics and smoking habits; (2) information that could be collected via mobile devices, including the time elapsed since the participant last smoked and the time of day; and (3) a unique participant identifier, which was incorporated as a categorical feature in the personalized models only. Including this identifier adds participant-specific parameters to the model, allowing predictions to be explicitly personalized. However, even when this identifier is omitted, the personalized model development scheme (ie, training on days 1-10 from all participants) allows

the model to learn from each participant's previously visited locations when predicting their current risk.

Mobile Device Implementation (QuitEye)

A nonpersonalized (out-of-sample) model incorporating image features only was implemented in TensorFlow Lite to allow prediction via mobile devices. Other features were omitted so that predictions could be made based on images only without additional data collection. A prototype mobile app was built using Flutter or Dart and tested on Google Pixel 3 (Android). QuitEye is applied to individual frames from a live video feed at a rate of approximately eight samples per second and is configured to display smoking and craving predictions corresponding to each frame.

Data Availability

The data sets analyzed in this study are not publicly available because they contain images of participants' personal daily environments that cannot be deidentified. However, the code supporting this work is available from the corresponding author upon reasonable request.

Results

Demographics and Descriptive Statistics

Of the 77 individuals screened for the study, 52 (68%) were eligible and consented to participate. Four participants were withdrawn or lost to follow-up, and the remaining 48 participants completed the study. One participant completed their study visits remotely because of in-person visit restrictions related to COVID-19. Among the participants who completed the study, a total of 8008 images were collected, 3648 (45.55%) of which were from completed random prompts and 4360 (54.45%) of which were from completed smoking prompts. Demographic characteristics, image details, and other descriptive statistics are presented in Table 1.

Table 1. Demographics and descriptive statistics (N=48).

Characteristics	Values
Demographics	
Sex	
Female:male	32:16
Female, n (%)	32 (67)
Age (years)	
Value, median (IQR)	40.5 (31-49)
Value, range	19-64
Race, n (%)	
White	31 (65)
Black or African American	19 (40)
American Indian	1 (2)
Native Hawaiian or Pacific Islander	1 (2)
Ethnicity, n (%)	
Not Hispanic or Latino	46 (96)
Hispanic or Latino	2 (4)
Smoking history	
Cigarettes per day (weekday)	
Value, median (IQR)	15 (12-20)
Value, range	7-30
Cigarettes per day (weekend)	
Value, median (IQR)	15 (14-20)
Value, range	10-30
Fagerstrom test of nicotine dependence	
Value, median (IQR)	6 (4-7)
Value, range	2-9
Images	
Total images taken	
Value, median (IQR)	163 (117-200)
Value, range	63-406
Images when smoking	
Value, median (IQR)	87 (67-132)
Value, range	25-322
Images when craving	
Value, median (IQR)	58 (19-99)
Value, range	1-210
Images when smoking permitted	
Value, median (IQR)	122 (96-160)
Value, range	25-388
Images when outside	
Value, median (IQR)	45 (20-78)
Value, range	3-183

Model Performance

Without personalization (out-of-sample performance), QuitEye predicted smoking with AUC=0.723 and average precision (AP)=0.785, craving with AUC=0.522 and AP=0.427, whether smoking was permitted with AUC=0.815 and AP=0.937, and whether the participant was outside with AUC=0.929 and AP=0.922. With personalization, performance was substantially improved: QuitEye predicted smoking with AUC=0.827 and AP=0.882, craving with AUC=0.837 and AP=0.789, whether smoking was permitted with AUC=0.932 and AP=0.981, and whether the participant was outside with AUC=0.977 and AP=0.956 (Figures 2 and 3).

Figure 2. Receiver operating characteristic curves for each of the four outcomes for both the nonpersonalized (out-of-sample) and personalized (longitudinal) models. AUC: area under the receiver operating characteristic curve.

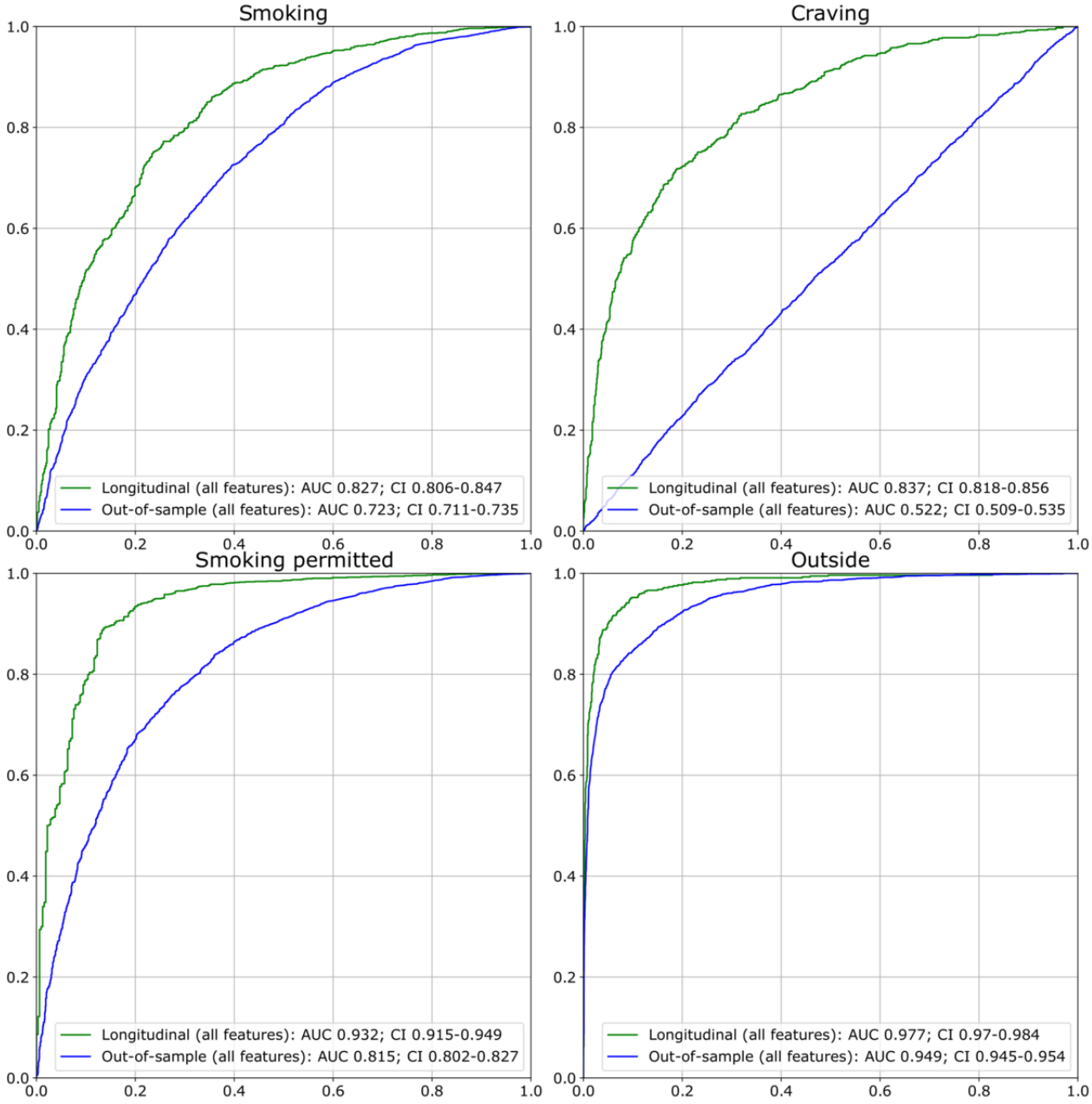


Figure 3. Precision recall curves for each of the four outcomes for both the nonpersonalized (out-of-sample) and personalized (longitudinal) models.

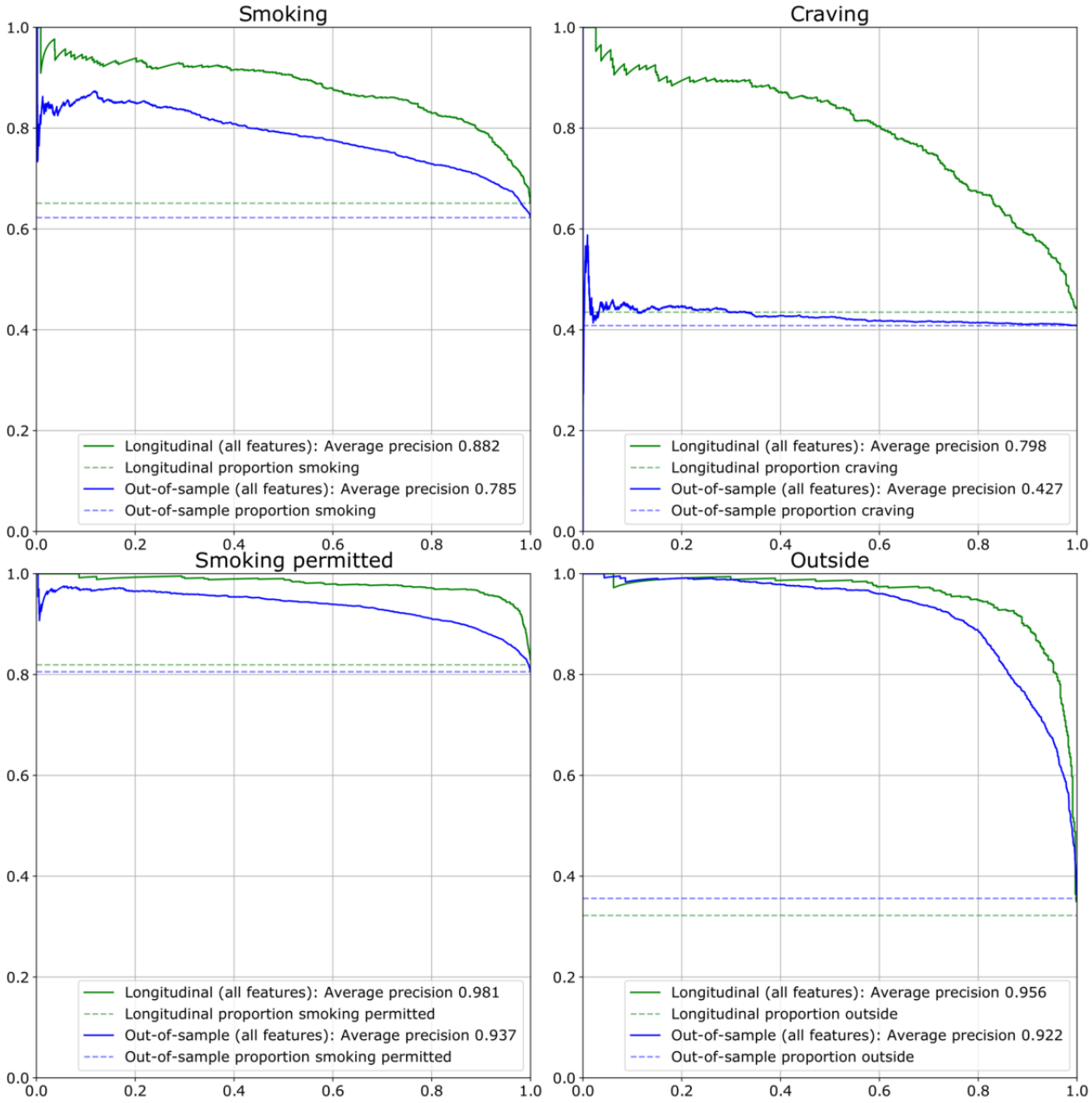


Image features were critical to these performance figures for all outcomes except craving. In the nonpersonalized (out-of-sample) models, removing the image features lowered AUC by 0.221 when predicting smoking, by 0.229 when predicting whether smoking was permitted, and by 0.253 when predicting whether the participant was outside but by only 0.027 when predicting craving. In the personalized (longitudinal) models, removing the image features lowered AUC by 0.192 when predicting smoking, by 0.168 when predicting whether smoking was permitted, and by 0.178 when predicting whether the participant was outside but increased AUC by 0.034 when predicting craving (Table 2).

In the out-of-sample models, baseline information about household smoking locations improved the prediction of craving (ΔAUC=0.050) and whether smoking was permitted (ΔAUC=0.020), but other nonimage features had less impact. Surprisingly, knowing the time since the last cigarette did not improve the prediction of smoking (ΔAUC=−0.014) or craving (ΔAUC=−0.026; Table 2). Performance among individual participants when predicting smoking was highly correlated with performance predicting whether smoking was permitted (Spearman ρ=0.55; P<.001), whereas correlations between other pairs of outcomes were not statistically significant.

In the personalized models, the participant identifier substantially improved the prediction of craving (ΔAUC=0.070), and baseline information about smoking locations outside of the household slightly improved the prediction of craving (ΔAUC=0.013); however, nonimage features had little effect on performance (ΔAUC<0.007). Similar to the out-of-sample models, performance among individual participants when predicting smoking was highly correlated with performance predicting whether smoking was permitted (r=0.71; Spearman ρ<0.001). However, this was not the case for any other pairs of outcomes.

Analyses of model calibration showed that outcome probabilities predicted by QuitEye were consistent with true outcome rates, except when predicting craving via the out-of-sample model (Figure 4). This suggests that the predicted smoking probability reflects the true environment-associated smoking probability.

Table 2. Model performance (area under the receiver operating characteristic curve) before and after removal of specific data elements.

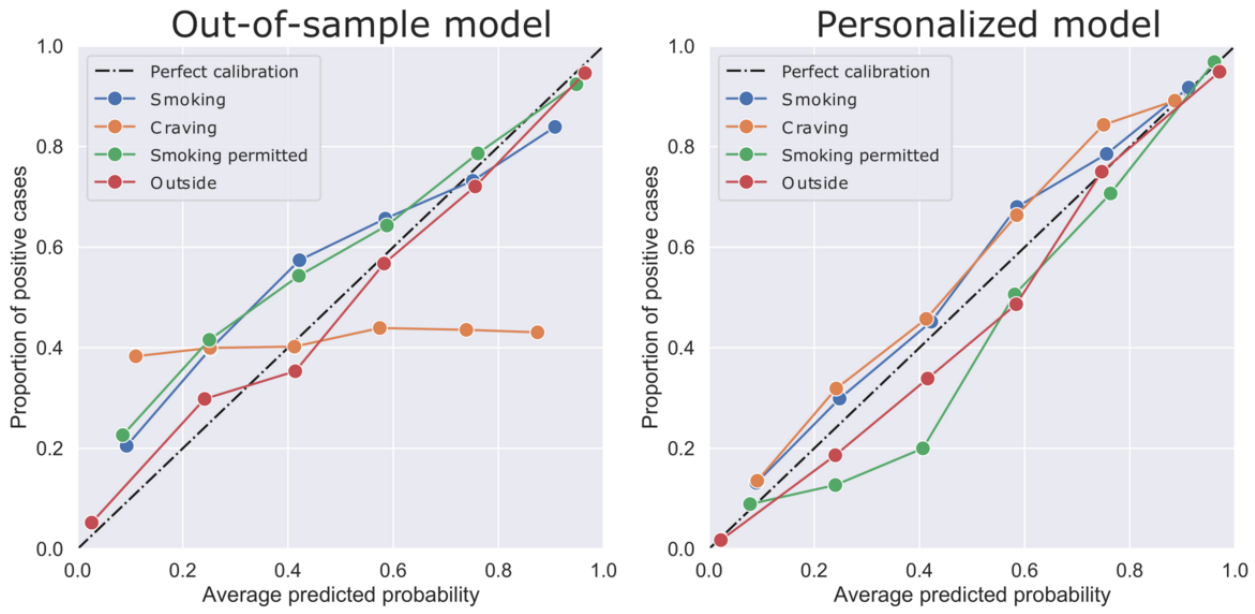
Model performance	Area under the receiver operating characteristic curve (Δ[a])			
	Smoking	Craving	Smoking permitted	Outside
Out-of-sample models				
Base model (all features)	0.723 (N/A[b])	0.522 (N/A)	0.815 (N/A)	0.949 (N/A)
Images	0.502 (−0.221)	0.495 (−0.027)	0.586 (−0.229)	0.696 (−0.253)
Demographics	0.729 (0.006)	0.542 (0.021)	0.810 (−0.005)	*0.952*[c] (0.002)
Time since last cigarette	*0.737* (0.014)	0.548 (0.026)	*0.819* (0.004)	0.944 (−0.005)
Time of day, weekday or weekend	0.726 (0.003)	*0.553* (0.032)	0.806 (−0.008)	0.945 (−0.004)
Household smoking locations	0.735 (0.012)	0.472 (−0.050)	0.795 (−0.020)	0.950 (0.001)
Other smoking locations	0.717 (−0.006)	0.513 (−0.009)	0.812 (−0.003)	0.948 (−0.001)
Longitudinal models				
Base model (all features)	0.827 (N/A)	0.837 (N/A)	0.932 (N/A)	*0.977* (N/A)
Images	0.635 (−0.192)	0.871 (0.034)	0.764 (−0.168)	0.799 (−0.178)
Demographics	0.824 (−0.002)	0.836 (−0.002)	0.929 (−0.003)	0.976 (−0.001)
Time since last cigarette	0.828 (0.002)	0.840 (0.003)	*0.938* (0.006)	0.975 (−0.002)
Time of day, weekday or weekend	*0.831* (0.004)	*0.844* (0.007)	0.929 (−0.003)	0.975 (−0.002)
Household smoking locations	0.826 (0.000)	0.836 (−0.002)	0.925 (−0.007)	0.976 (−0.001)
Other smoking locations	0.824 (−0.003)	0.824 (−0.013)	0.929 (−0.003)	0.976 (−0.001)
Personal identifier	0.829 (0.002)	0.767 (−0.070)	0.926 (−0.006)	0.975 (−0.002)

[a]Change in the area under the receiver operating characteristic curve compared to the base model.

[b]N/A: not applicable.

[c]Italics indicate the best performing model for that outcome.

Figure 4. Calibration curves for the nonpersonalized (out-of-sample; left panel) and personalized (longitudinal; right panel) models. Model-predicted probabilities are aggregated by percentile (N=6 bins), then compared with the proportion of positive outcomes in each bin. Good calibration implies that model predictions are an accurate estimate of the true probability of a positive outcome.



Effect of Location on Performance

Analyses of model performance by self-reported location type showed that QuitEye is more effective in some locations than others. For the nonpersonalized (out-of-sample) model (Figure 5), smoking prediction was most effective at work (AUC=0.848), followed by stores or restaurants (AUC=0.806). Across all prediction tasks, prediction tended to be less effective for images taken within vehicles. Many of these images showed outside scenery as viewed from the vehicle, leading the model to incorrectly predict that the participant was outside (AUC=0.603) rather than inside the vehicle.

Improvements in performance from personalized training also varied by location (Figures 5 and 6). AUC was improved most for outdoor home locations—prediction of smoking was improved by ΔAUC=0.230 in these locations, and prediction of whether smoking was permitted was improved by ΔAUC=0.435.

Smoking prediction was more effective for participants whose self-reported location type was more variable (r=0.48; *P*=.001), quantified as the entropy of self-reported location type. This effect was not observed in other prediction tasks. Smoking prediction was also more effective for those with higher mutual information between smoking and self-reported location type (r=0.53; *P*<.001); the mutual information quantifies the degree to which location type provides information about smoking behavior (Figure 7). This effect was not observed in other prediction tasks.

Figure 5. Outcomes and model performance by location type (out-of-sample). The bar plots indicate the proportion of positive outcomes (with SE) by self-reported location type, and the line plots indicate model performance (average precision) for images taken in each location. NA: prediction performance is not applicable, because there is no variability in the outcome in this location type.

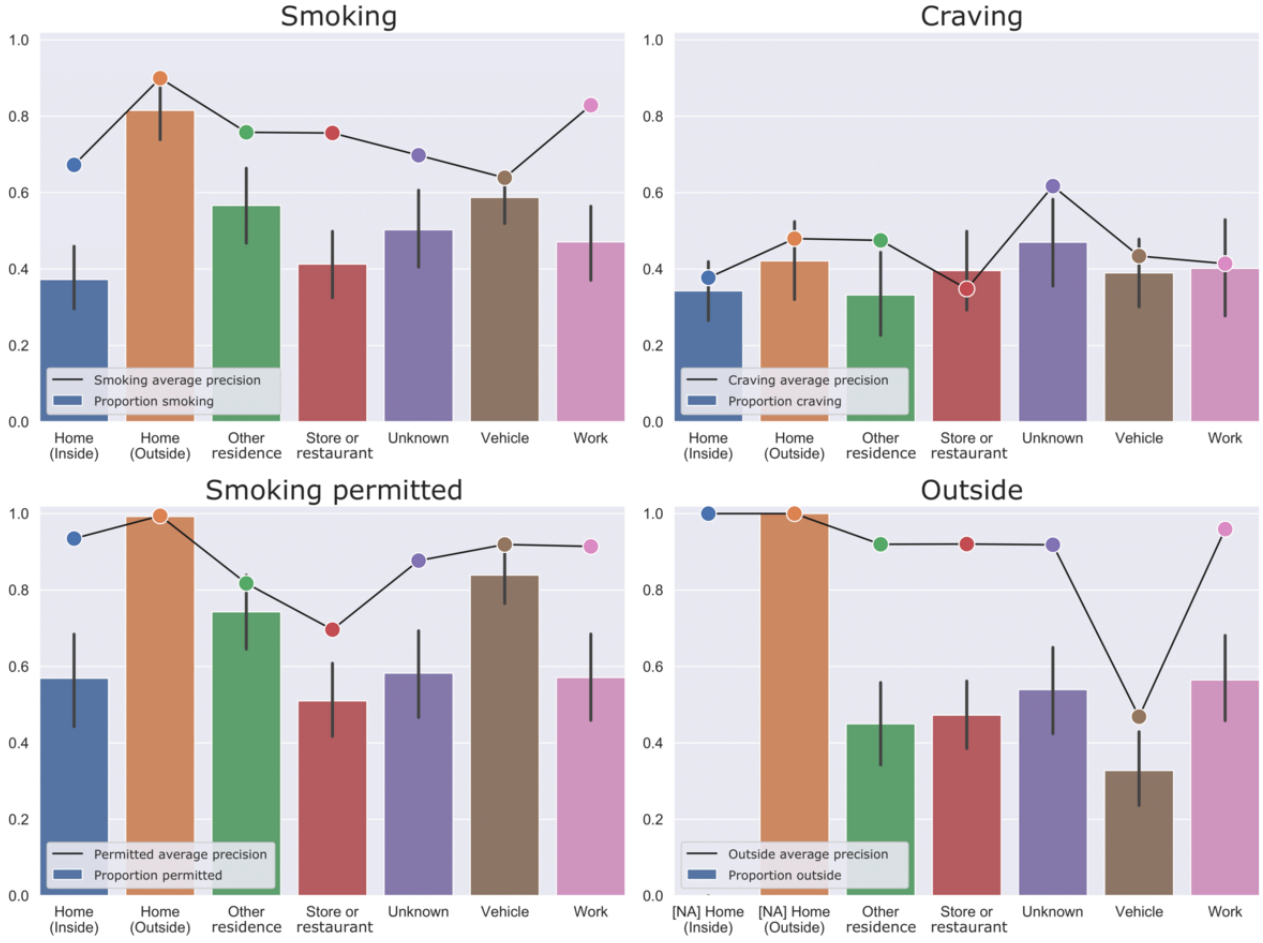


Figure 6. Outcomes and model performance by location type (personalized). The bar plots indicate the proportion of positive outcomes (with SE) by self-reported location type, and the line plots indicate model performance (average precision) for images taken in each location. NA: prediction performance is not applicable, because there is no variability in the outcome in this location type.

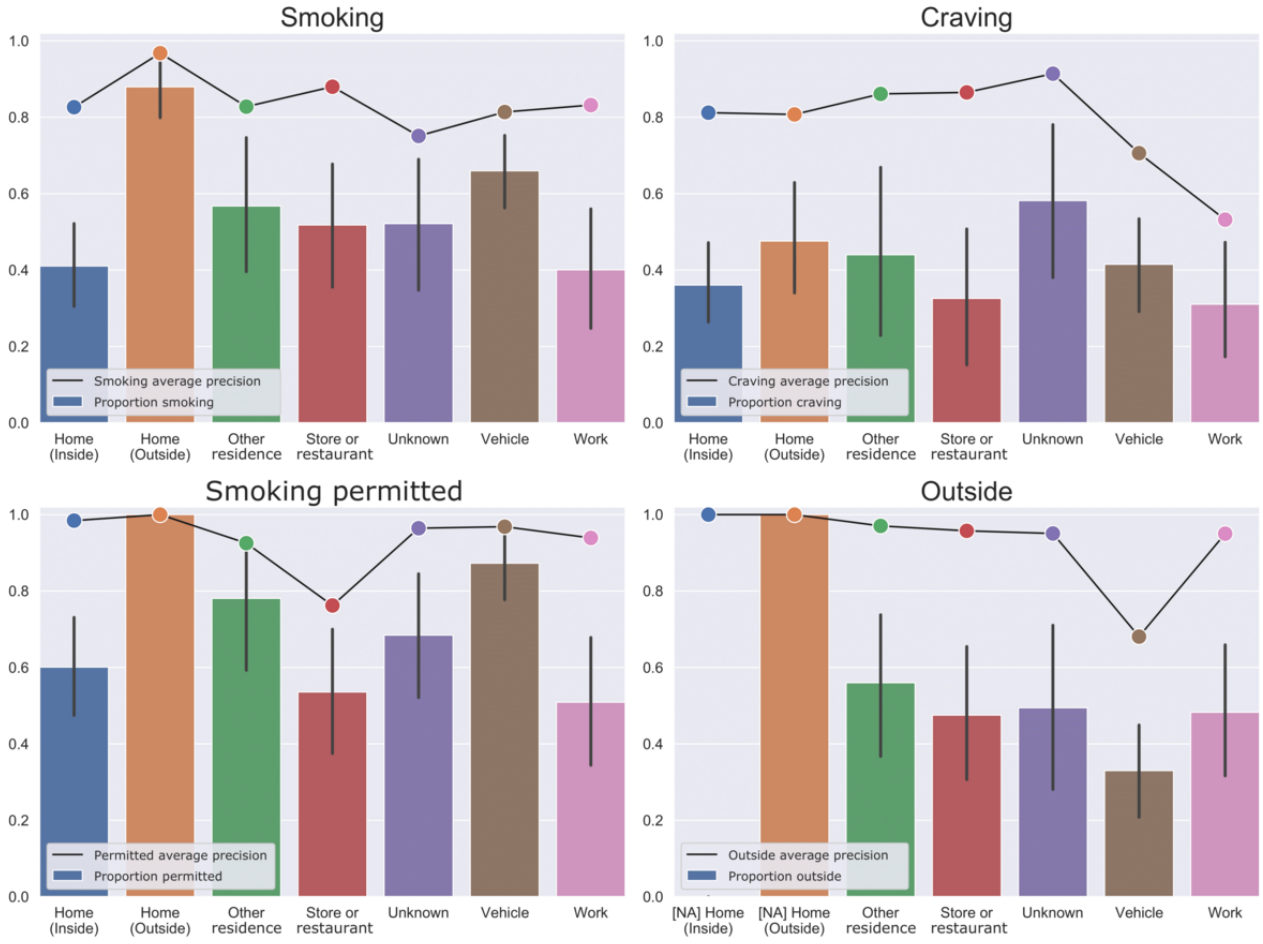
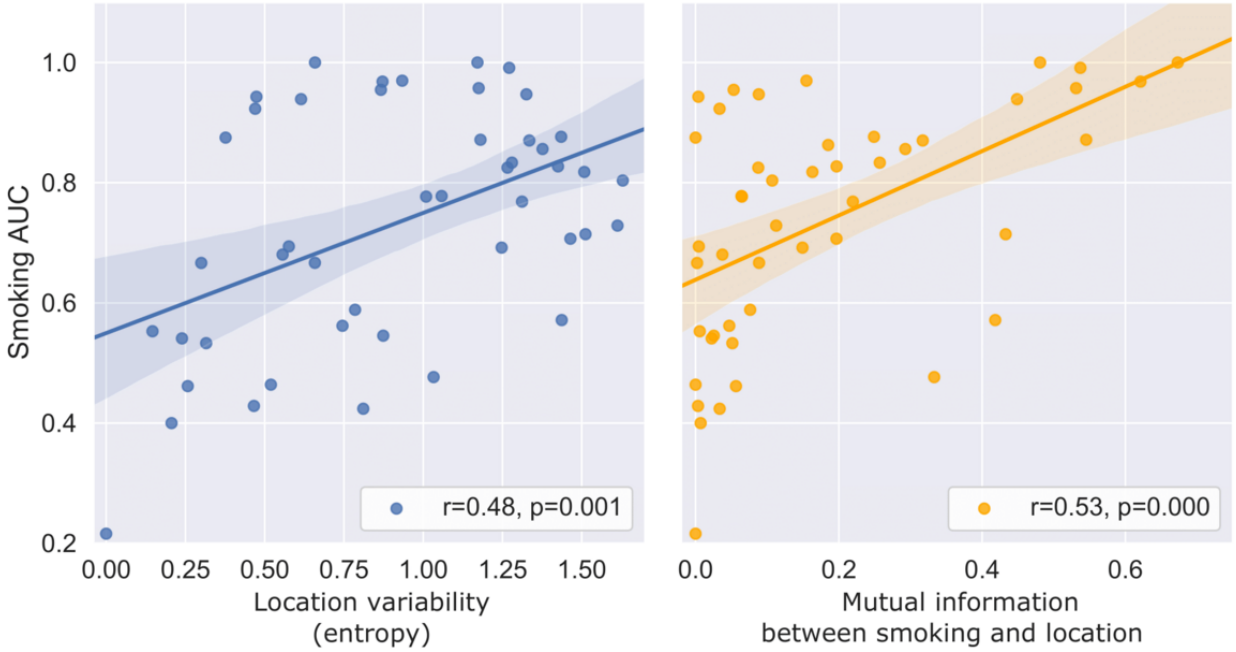


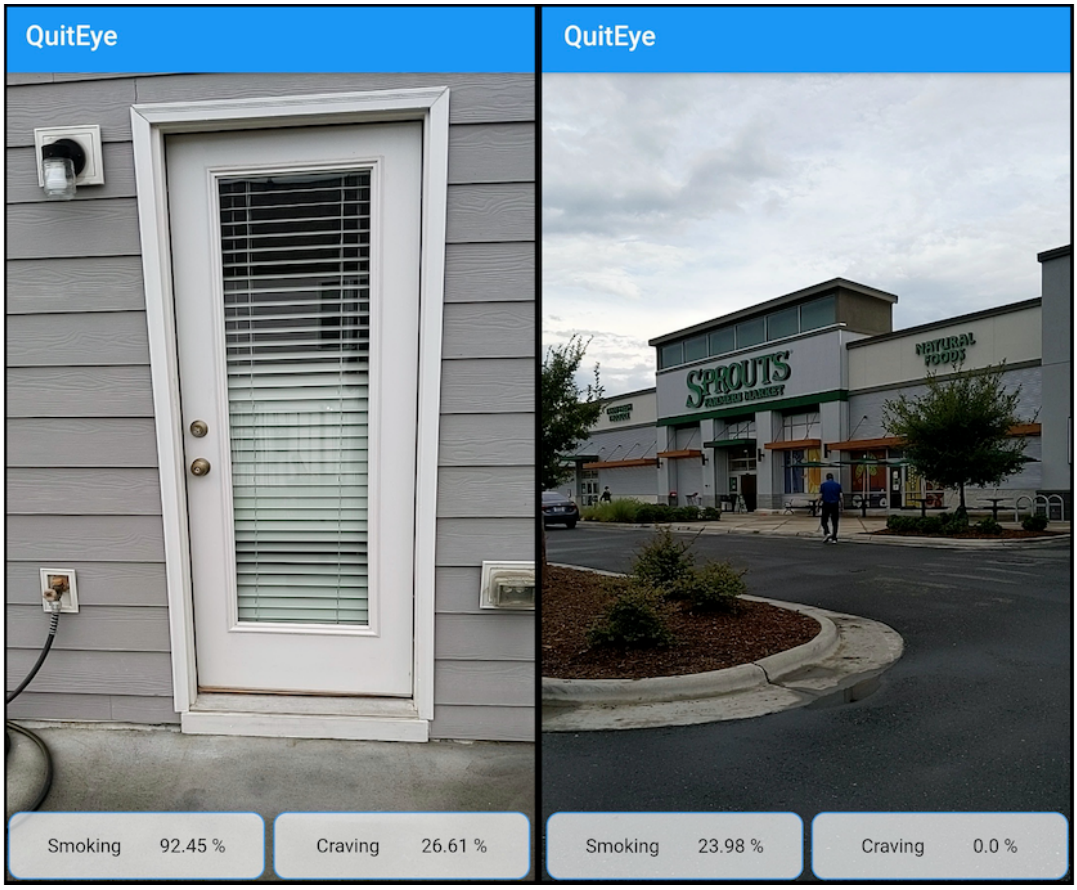
Figure 7. Effect of location variability on performance. Higher performance of smoking risk prediction among individual participants is associated with higher variability in self-reported locations (left panel) and higher mutual information between self-reported location type and smoking status (right panel). AUC: area under the receiver operating characteristic curve.



Mobile Implementation

Screenshots of real-time smoking and craving risk prediction using the QuitEye mobile app are presented in Figure 8.

Figure 8. Mobile implementation of QuitEye. Screenshots of real-time smoking and craving risk prediction via the QuitEye mobile app in a high smoking risk environment (left panel) and a low smoking risk environment (right panel).



Discussion

Principal Findings

A growing body of knowledge suggests that habitual smoking environments promote craving and smoking behaviors. Our previous study demonstrated that computer vision could distinguish between habitual smoking and nonsmoking environments, leading us to hypothesize that real-world smoking risk has important, quantifiable environmental correlates that can be leveraged to predict smoking behaviors more effectively in real time. This study confirms this hypothesis: QuitEye effectively predicted smoking status and associated outcomes across the full range of environments encountered by our sample of smokers in their daily lives. By learning from other smokers' behaviors, our models can scan a smoker's environment to predict their current smoking risk. QuitEye also predicted whether smoking was permitted in the current location and whether the smoker was inside or outside, providing important context relevant to smoking cessation interventions. The results show that knowledge of recent smoking and daily smoking habits (eg, time of day) improved these predictions, but it is the images themselves that contributed most to good prediction performance for all outcomes except craving.

Importantly, these results provide additional evidence that the environmental correlates of smoking vary among smokers. Our nonpersonalized models achieved good out-of-sample prediction performance, suggesting that environmental factors are shared among smokers more than they are distinct. However, model personalization led to statistically significant improvements in smoking risk prediction, implying that there are meaningful environmental correlates of smoking behaviors that are specific to individual smokers.

To achieve personalization in this study, smokers had to self-report their smoking behaviors for 10 days while collecting images of their daily environments. These data were then used to refine the prediction model. This process is burdensome but may be particularly important for smokers whose daily environments are atypical or whose smoking behaviors do not follow common patterns. If this is not possible, a lesser degree of personalization can be achieved by asking smokers to provide information about the locations where they commonly smoke. Alternatively, models can be iteratively improved during use, for example in a mobile app, by prompting the user to confirm or deny smoking predictions made by the model.

The ability to predict environment-associated smoking risk in real time unlocks a range of environment-focused smoking cessation interventions. Real-time risk prediction can be used to trigger a JITAI [9] in which support is provided to smokers via a mobile device at the time and place when it is most needed [31]. Information about environment-associated risk can be

combined with information about the smoker's internal state, provided by wearable devices, to obtain a more comprehensive estimate of smoking risk and craving. We anticipate that information about the external environment would enhance the prediction of smoking risk from other data streams (eg, heart rate from wearable devices), but the degree to which these data sources complement one another has yet to be explored. Ultimately, we envision a system in which wearable eyewear (eg, smart glasses) continuously acquires images of the user's external environment, alerts them with visual feedback when high-risk environments are detected, and suggests appropriate coping strategies for a given context. As augmented reality technologies mature [22], this paradigm can be used not only to support smoking cessation but also to understand and respond to environmental correlates of a broader range of health-related symptoms and behaviors.

However, several other intervention types are possible. For example, QuitEye could be used to identify environmental correlates of smoking risk in a smoker's daily environment before a quit attempt, allowing them to restructure their environments or daily activity patterns to increase the likelihood of quitting successfully. As QuitEye can predict the smoking risk associated with any image, including images of locations not yet visited, it can help smokers preemptively avoid visiting prosmoking environments. During a quit attempt, for instance, a smoker might choose to visit a restaurant that has a lower smoking risk, as determined based on images available on the internet.

Although images improved the prediction of smoking substantially, they did not improve the prediction of craving. To the contrary, our best-performing craving prediction models do not incorporate image features, and out-of-sample prediction performance for craving was poor. Our laboratory research suggests that habitual smoking environments do provoke craving, but this study's results do not provide additional support for this finding. Consequently, the role of environmental factors in the emergence of craving, or in the progression from craving to smoking itself, remains unclear. These conflicting findings may be partly owing to our EMA procedure. At the time of smoking, participants were asked to report their craving *before* smoking. Thus, the corresponding image taken at the time of smoking may not match the external environment corresponding to self-reported craving. However, in a follow-up analysis, we attempted to predict craving only from random prompts, and similar AUC values were observed. Further studies are needed to examine the environmental correlates of craving more thoroughly. In particular, continuous image acquisition (via wearable cameras or smart glasses) may provide additional environmental or social cues that are relevant to craving but not captured by a single image. Alternatively, general craving may be driven more by internal (eg, low plasma nicotine levels and negative affect) rather than external factors.

As shown in Figure 7, smoking risk prediction was more effective for participants whose locations were more variable and for whom location type provided more information about smoking behavior. Some participants' smoking environments were mostly distinct from their other daily environments, allowing our models to identify a strong relationship between environmental factors and smoking status. Other participants tended to smoke in the same environments where they spent most of their time, making it difficult to identify robust environment-smoking associations. Notably, this difference was not associated with demographic factors in this study. More studies are needed to understand these groups and determine whether there are identifiable subpopulations of smokers among whom QuitEye is particularly effective or ineffective.

Limitations

Although EMA provides more accurate smoking tracking than other self-report methods [32], smoking events may have been omitted or incorrectly reported. Our models predicted self-reported smoking, which may differ from true smoking events. Although our EMA procedure was designed to be brief and minimally burdensome, picture-taking and other EMA requirements may have increased the number of smoking events omitted by our participants. Furthermore, some EMA responses may not have been completed promptly upon smoking, thus reducing self-report validity. Other outcomes were also self-reported and subject to participant error or omission. As previously discussed, our EMA asked participants to self-report their craving *before* smoking in the smoking-initiated prompts, which may have introduced additional errors or variability.

In addition to smoking-initiated prompts, participants completed a total of six system-initiated prompts at randomly selected times throughout the day. More frequent prompts would have provided a more comprehensive sample of participants' daily environments, but this might have also resulted in reduced EMA adherence. Camera design and image quality varied among participants, who used their own smartphones to take pictures. Variability in image quality can be reduced by acquiring images using wearable cameras or smart glasses. This approach would also allow images to be captured throughout the day, providing complete information about the participants' daily environments.

Future Directions

This study did not include a quit attempt. The results showed that QuitEye predicts smoking risk effectively outside of a quit attempt, but its ability to predict lapses and relapse after quitting is unknown. Other (nonimage) data streams from mobile devices have been used to predict lapse risk [21], and we anticipate that QuitEye would provide complementary information about environment-associated risk factors. In future work, we will explore the relationship between the environmental correlates of smoking before quitting and the environmental correlates of lapse and relapse.

Owing to privacy concerns, participants were asked to avoid taking pictures of other people. However, this restriction may have prevented us from identifying interpersonal triggers and other important social determinants of smoking. The ability to recognize these and other dynamic environmental features is an important advantage of our approach compared with other sources of environmental information, such as GPS. In future work, we hope to explore the use of computer vision to identify the social determinants of smoking.

Now that QuitEye has been implemented as a mobile app, we can prospectively evaluate the real-time prediction of

environment-associated risk and develop *environment-aware* mobile health cessation interventions. These interventions will incorporate information about the environment in addition to user physiology and other information with recognized predictive values [11-15]. An important goal of this study is to quantify the contribution of each data source (eg, physiology vs environment) to overall prediction performance. Initial interventions incorporating QuitEye will use smartphone cameras to identify environment-associated smoking risks in smokers' daily environments. Later interventions will use smart eyewear to continuously acquire images of the smoker's environment and provide support when high-risk environments are encountered. An important component of this study will be to evaluate the benefit of just-in-time support versus the cost of false alarms, which will allow us to select an appropriate risk threshold for triggering the intervention.

Conclusions

Images of daily environments can be used to predict smoking risk effectively. Our risk prediction system, QuitEye, also predicts craving, whether smoking is permitted, and whether the participant is outside, providing important contextual information that could inform JITAIs for smoking cessation. Performance can be further improved through personalization, achieved by (1) fine-tuning QuitEye with images of a given smoker's daily environment or (2) asking participants to provide information about their habitual smoking environments. QuitEye has been optimized for mobile devices and implemented as a mobile app, allowing environment-associated smoking risk to be continuously assessed in a mobile device-based smoking cessation intervention.

Acknowledgments

Funding support for this study was provided by National Institute on Drug Abuse R21DA047131 (FJM) and K23DA042898 (JAO). The authors would like to thank Christianne Carson, Anthony DeVito, and Patricia Sabo for their assistance with data collection and processing, Abhishek Jadhav for assistance with preliminary analyses, and Cynthia Conklin for conceptual support and feedback.

Authors' Contributions

MME, JAO, and FJM developed the concept. MME, JAO, RK, and FJM designed the study. RK and FJM oversaw this study. MME, JAO, JD'A, and FJM designed the analyses. MME and JD'A conducted the analyses. JD'A developed the mobile app. MME drafted the manuscript. All authors have revised the manuscript accordingly.

Conflicts of Interest

MME, JAO, and FJM declare that they are authors of a US patent app related to this work and have no competing interests. JD'A and RK declare that they have no competing interests.

References

1. US Department of Health and Human Services. The health consequences of smoking—50 years of progress. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta. 2014. URL: https://aahb.org/Resources/Pictures/Meetings/2014-Charleston/PPT%20Presentations/Sunday%20Welcome/Abrams.AAHB.3.13.v1.o.pdf [accessed 2020-08-11]
2. Tobacco. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2020-08-11]
3. Walton K, Wang TW, Schauer GL, Hu S, McGruder HF, Jamal A, et al. State-specific prevalence of quit attempts among adult cigarette smokers - United States, 2011-2017. MMWR Morb Mortal Wkly Rep 2019 Jul 19;68(28):621-626 [FREE Full text] [doi: 10.15585/mmwr.mm6828a1] [Medline: 31318853]
4. Smoking cessation: a report of the Surgeon General. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/data_statistics/sgr/2020-smoking-cessation/index.html [accessed 2021-09-01]
5. 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]
6. Pallejà-Millán M, Rey-Reñones C, Uriarte ML, Granado-Font E, Basora J, Flores-Mateo G, et al. Evaluation of the Tobbstop mobile app for smoking cessation: cluster randomized controlled clinical trial. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e15951 [FREE Full text] [doi: 10.2196/15951] [Medline: 32589153]
7. Regmi K, Kassim N, Ahmad N, Tuah N. Effectiveness of mobile apps for smoking cessation: a review. Tob Prev Cessat 2017 Apr 12;3:12 [FREE Full text] [doi: 10.18332/tpc/70088] [Medline: 32432186]
8. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]

9. 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]
10. 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]
11. Sharmin M, Weber T, Sarker H, Saleheen N, Kumar S, Ahmed S, et al. Opportunities and challenges in designing participant-centric smoking cessation system. In: Proceedings of the IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); Jul 4-8, 2017; Turin, Italy. [doi: 10.1109/compsac.2017.184]
12. Zhai D, Schiavone G, Van Diest I, Vrieze E, DeRaedt W, Van Hoof C. Ambulatory Smoking Habits Investigation based on Physiology and Context (ASSIST) using wearable sensors and mobile phones: protocol for an observational study. BMJ Open 2019 Sep 05;9(9):e028284 [FREE Full text] [doi: 10.1136/bmjopen-2018-028284] [Medline: 31492781]
13. Imtiaz M, Senyurek V, Belsare P, Tiffany S, Sazonov E. Objective detection of cigarette smoking from physiological sensor signals. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019 Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8856831]
14. Abo-Tabik M, Costen N, Darby J, Benn Y. Towards a smart smoking cessation app: a 1D-CNN model predicting smoking events. Sensors (Basel) 2020 Feb 17;20(4):1099 [FREE Full text] [doi: 10.3390/s20041099] [Medline: 32079359]
15. Engelhard M, Xu H, Carin L, Oliver JA, Hallyburton M, McClernon FJ. Predicting smoking events with a time-varying semi-parametric hawkes process model. In: Proceedings of the Machine Learning for Healthcare Conference. 2018 Presented at: Proceedings of the Machine Learning for Healthcare Conference; Aug 17-18, 2018; Palo Alto, California URL: https://proceedings.mlr.press/v85/engelhard18a.html
16. Conklin CA, Perkins KA, Robin N, McClernon FJ, Salkeld RP. Bringing the real world into the laboratory: personal smoking and nonsmoking environments. Drug Alcohol Depend 2010 Sep 01;111(1-2):58-63 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.03.017] [Medline: 20510552]
17. Ahnallen CG, Tidey JW. Personalized smoking environment cue reactivity in smokers with schizophrenia and controls: a pilot study. Psychiatry Res 2011 Jul 30;188(2):286-288 [FREE Full text] [doi: 10.1016/j.psychres.2011.04.005] [Medline: 21550670]
18. Stevenson JG, Oliver JA, Hallyburton MB, Sweitzer MM, Conklin CA, McClernon FJ. Smoking environment cues reduce ability to resist smoking as measured by a delay to smoking task. Addict Behav 2017 Apr;67:49-52 [FREE Full text] [doi: 10.1016/j.addbeh.2016.12.007] [Medline: 28012965]
19. Conklin C, McClernon F, Vella E, Joyce C, Salkeld R, Parzynski C, et al. Combined smoking cues enhance reactivity and predict immediate subsequent smoking. Nicotine Tob Res 2019 Jan 04;21(2):241-248 [FREE Full text] [doi: 10.1093/ntr/nty009] [Medline: 29370401]
20. McClernon FJ, Conklin CA, Kozink RV, Adcock RA, Sweitzer MM, Addicott MA, et al. Hippocampal and insular response to smoking-related environments: neuroimaging evidence for drug-context effects in nicotine dependence. Neuropsychopharmacology 2016 Feb;41(3):877-885 [FREE Full text] [doi: 10.1038/npp.2015.214] [Medline: 26179147]
21. Suchting R, Hébert ET, Ma P, Kendzor DE, Businelle MS. Using elastic net penalized cox proportional hazards regression to identify predictors of imminent smoking lapse. Nicotine Tob Res 2019 Jan 04;21(2):173-179 [FREE Full text] [doi: 10.1093/ntr/ntx201] [Medline: 29059349]
22. Engelhard M, Oliver J, McClernon F. Digital envirotyping: quantifying environmental determinants of health and behavior. NPJ Digit Med 2020 Mar 12;3:36 [FREE Full text] [doi: 10.1038/s41746-020-0245-3] [Medline: 32195371]
23. Doherty AR, Hodges SE, King AC, Smeaton AF, Berry E, Moulin CJ, et al. Wearable cameras in health: the state of the art and future possibilities. Am J Prev Med 2013 Mar;44(3):320-323. [doi: 10.1016/j.amepre.2012.11.008] [Medline: 23415132]
24. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, et al. An ethical framework for automated, wearable cameras in health behavior research. Am J Prev Med 2013 Mar;44(3):314-319. [doi: 10.1016/j.amepre.2012.11.006] [Medline: 23415131]
25. Wilson G, Jones D, Schofield P, Martin DJ. Experiences of using a wearable camera to record activity, participation and health-related behaviours: qualitative reflections of using the Sensecam. Digit Health 2016 Nov 1;2:2055207616682628 [FREE Full text] [doi: 10.1177/2055207616682628] [Medline: 29942578]
26. Imtiaz M, Hossain D, Senyurek V, Belsare P, Tiffany S, Sazonov E. Wearable egocentric camera as a monitoring tool of free-living cigarette smoking: a feasibility study. Nicotine Tob Res 2020 Oct 08;22(10):1883-1890 [FREE Full text] [doi: 10.1093/ntr/ntz208] [Medline: 31693162]
27. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings ofthe IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT, USA p. 4510-4520. [doi: 10.1109/CVPR.2018.00474]

28. Engelhard MM, Oliver JA, Henao R, Hallyburton M, Carin LE, Conklin C, et al. Identifying smoking environments from images of daily life with deep learning. JAMA Netw Open 2019 Aug 02;2(8):e197939 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.7939] [Medline: 31373647]
29. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015 Apr 11;115(3):211-252. [doi: 10.1007/s11263-015-0816-y]
30. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014 Mar 29;6(1):10 [FREE Full text] [doi: 10.1186/1758-2946-6-10] [Medline: 24678909]
31. McClernon FJ, Choudhury RR. I am your smartphone, and I know you are about to smoke: the application of mobile sensing and computing approaches to smoking research and treatment. Nicotine Tob Res 2013 Oct;15(10):1651-1654 [FREE Full text] [doi: 10.1093/ntr/ntt054] [Medline: 23703731]
32. Shiffman S. How many cigarettes did you smoke? Assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment. Health Psychol 2009 Sep;28(5):519-526 [FREE Full text] [doi: 10.1037/a0015197] [Medline: 19751076]

Abbreviations

AP: average precision
AUC: area under the receiver operating characteristic curve
EMA: ecological momentary assessment
JITAI: just-in-time adaptive intervention
photoEMA: photograph-augmented ecological momentary assessment

Edited by R Kukafka; submitted 11.02.21; peer-reviewed by R Suchting, D Gunasekeran; comments to author 14.03.21; revised version received 01.04.21; accepted 01.08.21; published 01.11.21.

Please cite as:
Engelhard MM, D'Arcy J, Oliver JA, Kozink R, McClernon FJ
Prediction of Smoking Risk From Repeated Sampling of Environmental Images: Model Validation
J Med Internet Res 2021;23(11):e27875
URL: https://www.jmir.org/2021/11/e27875
doi:10.2196/27875
PMID:34723819

Original Paper

Benefits, Problems, and Potential Improvements in a Nationwide Patient Portal: Cross-sectional Survey of Pharmacy Customers' Experiences

Maria Sääskilahti[1], MSc; Anna Ojanen[1,2], MSc; Riitta Ahonen[1], PhD; Johanna Timonen[1], PhD

[1]School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

[2]Humalisto Pharmacy, Turku, Finland

Corresponding Author:
Maria Sääskilahti, MSc
School of Pharmacy
Faculty of Health Sciences
University of Eastern Finland
PO Box 1627
Kuopio, 70211
Finland
Phone: 358 403552505
Email: maria.saaskilahti@uef.fi

Abstract

Background: Patient engagement is a worldwide trend in health care. Patient portals have the potential to increase patients' knowledge about their health and care and therefore enhance patient engagement. Portal users' experiences are needed to determine if these portals work appropriately and if there are barriers to achieving the aims that were set before their implementation.

Objective: The aim of this study is to analyze pharmacy customers' experiences of the Finnish nationwide patient portal My Kanta in terms of benefits, problems, and potential improvements.

Methods: A questionnaire survey was conducted among pharmacy customers in the spring of 2019. The questionnaires (N=2866) were distributed from 18 community pharmacies across mainland Finland to customers aged ≥18 years who were purchasing prescription medicines for themselves or their children aged <18 years. Using open-ended questions, customers were asked about their experiences of the benefits and problems of My Kanta and what improvements could be made. Their responses were encoded and categorized using inductive content analysis, stored in SPSS Statistics for Windows, and analyzed using frequencies.

Results: Of the 2866 questionnaires, a total of 994 (34.68%) questionnaires were included in the analysis. Most respondents were My Kanta users (820/994, 82.5%); of these 820 users, 667 (81.3%) reported at least one benefit, 311 (37.9%) reported at least one problem, and 327 (39.9%) reported at least one potential improvement when using My Kanta. The most commonly mentioned benefits were opportunities to view health data (290/667, 43.5%) and prescriptions (247/667, 37%) and to renew prescriptions (220/667, 33%). The most extensively reported problems with My Kanta were that the portal lacks health data (71/311, 22.8%), navigating the service and searching for information is difficult (68/311, 21.9%), and the delay before health data are incorporated into the service (41/311, 13.2%). The most frequently suggested potential improvements were that My Kanta needs more comprehensive health data (89/327, 27.2%); the service should be easier to navigate and information easier to access (71/327, 21.7%); the service should have more functions (51/327, 15.6%); and health data should be entered into the portal more promptly (47/327, 14.4%).

Conclusions: Pharmacy customers reported more benefits than problems or potential improvements regarding the use of My Kanta. The service is useful for viewing health data and prescriptions and for renewing prescriptions. However, portal users would like to see more data and functions available in the portal and data searches to be made easier. These improvements could make the data and functions provided by the portal easier to view and use and hence promote patient engagement.

(J Med Internet Res 2021;23(11):e31483) doi:10.2196/31483

KEYWORDS

benefit; problem; improvement need; patient portal; patient engagement; experience; survey

Introduction

Background

Promoting patient engagement is a worldwide trend, with the goal of improving patient-centered care and providing greater safety in health care [1]. The opportunity to access one's own medical data is regarded as a key element in patient engagement, as it provides patients with more information about their own health and care. In recent decades, many countries have developed national eHealth strategies, including the implementation of electronic health record systems [2]. Some of these systems include an opportunity for patients to view their electronic health records via patient portals. However, enhancing patient engagement requires not only implementing patient portals for data viewing but also ensuring that these services are user-friendly and can be used appropriately [1]. Collecting data regarding patients' perceptions, needs, and experiences when updating patient-centered health services is also part of patient engagement. In the case of patient portals, it is important to understand how patients use the medical data provided and whether the portal has achieved the goals set before its implementation. The aims of portals are to increase patients' empowerment and thus encourage them to take more responsibility for their own health and care.

As patient portals have only been introduced in recent years, only a limited number of studies have investigated users' experiences of their usability. However, some studies have been conducted, mainly in the United States and the Netherlands, where patient portals are targeted at certain regions, diseases, or organizations [3-7]. On the other hand, few studies have investigated nationwide patient portals [8-10], partly because they have been implemented in only a few countries, mainly Nordic countries. Studies focused on users' experiences of patient portals have shown that users are mostly satisfied with these portals [6-10]. Commonly reported benefits that portals have provided for their users are that medical records are viewable and can be read before or after visits and that users find themselves better informed and more empowered and engaged in their own care [3-7,9-11]. Being able to communicate with health care professionals has also been viewed as beneficial [3-6,11]. The problems encountered when using portals are mostly related to difficulty in understanding the information recorded within the portal and to the log-in process [3-5,10,11]. Users have reported that the portals could be improved by including more information, facilitating faster entry of data into the portals, and making the information more understandable [4,8-11]. Previous studies seem to indicate that experiences of patient portal use are fairly similar regardless of whether the portal is nationwide or specific to a certain area, disease, or organization.

Study Context

Kanta Services is a Finnish nationwide entity providing services such as Patient Data Repository and Prescription Center into which health records and prescriptions recorded in all public and private health care units and pharmacies across Finland are saved [12]. Users of the services are health care professionals, pharmacists, and the public.

My Kanta, a nationwide patient portal, is part of Kanta Services and allows patients to access their health and prescription information [13]. Access to My Kanta requires a Finnish social security number and electronic authentication, such as a web banking code, an electronic ID, or a mobile certificate. My Kanta was initially introduced in 2010, and its content has been extended over the years [12,13]. In its early years, it only provided an opportunity to view e-prescriptions. Today, however, patients can browse their own or their dependents' e-prescriptions and health data recorded as part of both public and private health care (Textboxes 1 and 2) [13,14,15]. Patients can also request prescription renewal and save their living will and organ donation testament. My Kanta also provides an opportunity to view in which health care units or pharmacies the user's own data have been browsed or processed. We have reported a more detailed description of My Kanta and the Finnish e-prescription system in a previous study [16].

Textbox 1. Contents of My Kanta.

Contents

- e-Prescriptions (all prescriptions have been issued electronically in Finland since 2017)
  - Name, dosage, and indication of medicine
  - Prescribing date and organization, prescriber
  - Valid date of prescription
  - Quantities of medicines outstanding
  - Medicine purchases and dispensing events
  - Prescription renewals
  - Health care units and pharmacies where prescription information has been processed
- Health data
  - Records of health care visits, diagnoses, critical risk factors
  - Laboratory test results, x-ray examinations
  - Referrals, medical certificates and reports
  - Health care units where health records have been browsed

Textbox 2. Functions of My Kanta.

Functions

- Viewing e-prescriptions and health data
- Printing out a summary of e-prescriptions
- Requesting a prescription renewal
- Giving informed consent to share health data
- Preventing the sharing of certain health data or e-prescriptions
- Saving a living will and organ donation testament
- Giving consent to other European pharmacies to dispense e-prescriptions (currently possible in Estonia, Croatia, and Portugal)
- Acting on behalf of dependents (at the time of the study, this refers to dependents aged <10 years. Access has only been extended gradually to underage dependents, aged 10-17 years, since October 2020).
  - Viewing e-prescriptions and health data
  - Submitting a prescription renewal request
  - Giving informed consent to share a dependent's health data

According to Kanta Services statistics, the My Kanta portal is widely used in Finland [17]. In 2019, more than 2 million people (total Finnish population 5.5 million) were using My Kanta, and the total number of log-ins was about 21 million. The familiarity of pharmacy customers with My Kanta, portal use, and experiences of its usability for viewing e-prescriptions were investigated in its early years in 2015 [18]. Since then, health data have been added to the portal, and portal use has increased significantly; hence, up-to-date studies are now needed. This study is part of a larger research project studying My Kanta use for viewing and monitoring health and prescription information from the patients' perspective. We have previously reported on the use of different functions and the usability of the service surveyed by means of structured questions and have investigated the factors related to the use and nonuse of the service [16,19].

Objectives

The aim of this study is to investigate pharmacy customers' experiences of the benefits and problems of the Finnish nationwide patient portal My Kanta and how it could be improved.

Methods

Study Setting

A questionnaire survey was conducted in early spring 2019 of pharmacy customers aged ≥18 years who were purchasing prescription medicines for themselves or their children <18 years. This target group was chosen as we wanted to reach people who potentially have a need to use My Kanta. Questionnaires (Multimedia Appendix 1) were distributed in

18 community pharmacies from 6 regions across mainland Finland. Using convenience sampling, 1 university pharmacy branch (owned by a university but operating as a privately owned pharmacy), 1 large privately owned city pharmacy, and 1 small privately owned rural pharmacy were recruited from each region. The number of questionnaires delivered to each pharmacy was in proportion to the number of prescriptions dispensed annually by the pharmacy and ranged between 40 and 320. A total of 3560 questionnaires were mailed to the pharmacies. Pharmacists were instructed to inform all eligible customers about the survey after dispensing prescription medicines and to offer them the questionnaire. Pharmacies were not required to keep a record of customers who declined to participate. Customers filled in the questionnaires at home and mailed them in return envelopes to the research group. Questionnaires were handed out as long as there were forms left for a maximum of 2 weeks. After the study period, pharmacists reported the number of remaining questionnaires to allow the calculation of the response rate. Reminders could not be sent because the respondents' personal data were not collected. Altogether, 2866 questionnaires were distributed.

Questionnaire

The 4-page questionnaire included 22 structured, Likert-scale, and open-ended questions (Multimedia Appendix 1). The questionnaire consisted of 3 parts. The first part was for all respondents and concerned background information, the second part was for those respondents who used My Kanta, and the third part was for respondents who did not use My Kanta. The questionnaire was designed using My Kanta pages and previous surveys of patient portals [7,13,18,20-25]. It was initially tested for face validity by 3 researchers who were experienced in designing questionnaire surveys. Thereafter, the questionnaire and data collection procedure were piloted in a pharmacy, with minor revisions made as a result.

This study reports the results from 3 open-ended questions from the second part of the questionnaire. The questions were as follows: *What advantages or benefits has the use of My Kanta provided to you?What problems have you experienced when using My Kanta*? *How could My Kanta be improved to make it easier for you to monitor and manage your medication and health information?* We used open-ended questions to ask about the benefits, problems, and potential improvements to get an overview of issues that spontaneously come to respondents' minds when they think about their experiences of My Kanta use. Background information was obtained by means of structured questions, except for the respondent's year of birth and number of regularly used prescription medicines, which were obtained using open-ended questions.

Data Analysis

A 2-phase analysis (qualitative and quantitative) was used. In the qualitative analysis, responses to open-ended questions were encoded and categorized using inductive content analysis. The analysis started by recording the answers in a table in Word 2016 (Microsoft Corp). This was continued to the point of saturation, which means that no new aspects related to the research questions emerged in the answers. After the saturation point, the remaining answers were examined, and only supplements to the previous aspects and new ideas were recorded. The saturation points for benefits, problems, and potential improvements were questionnaires 176, 248, and 250, respectively. In addition, if the answer to the question about benefits included a problem, it was moved to the analysis of problems, and if the answer to problems included an improvement idea, it was moved to the analysis of potential improvements. An analysis unit could be a single word, a sentence, or a group of sentences describing an idea related to benefits, problems, or potential improvements. Accordingly, an answer pertaining to more than 1 subject was separated into several analysis units. Simplifications were then made using these units. The simplifications were compared and then sorted into emerging subcategories, which were named according to all the simplifications in that subcategory. Similar subcategories were combined into main categories, and the main categories were named according to their content. Each questionnaire was then studied, and the responses were encoded into the main categories formed. The categorized data were stored in SPSS Statistics for Windows 10 (version 27.0; IBM Corp) for quantitative analysis. Data were analyzed using frequencies. Inductive content analysis was conducted by 2 researchers (AO and MS) after the simplifications were made. Contradictory categorizations were discussed by the research group. Content analysis was discussed by the research group throughout the process.

Ethical Statement

Ethical approval required by the funding organization was granted by the Committee on Research Ethics of the University of Eastern Finland (statement 23/2018). Participation in the survey was voluntary. Filling in the questionnaire and mailing it to the research group were regarded as informed consent to participate. No incentives were provided to the participants. Pharmacy owners consented to handing out questionnaires in their pharmacies.

Results

Study Population

Altogether, 996 questionnaires were returned. Two questionnaires, however, were blank and were therefore excluded. Consequently, 34.68% (994/2866) of questionnaires were included in the study. Most respondents were female (687/994, 69.4%; Table 1). The mean age of the respondents was 62 (SD 14.484; range 18-99) years. Most respondents (820/994, 82.5%) were My Kanta users. The characteristics of My Kanta users were very similar to those of all respondents, with the exception of more frequent internet use and internet use to search for health-related information.

Table 1. Characteristics of the questionnaire respondents.

Characteristics	Total respondents (n=994), n (%)	My Kanta users (n=820), n (%)
Gender[a]	990 (99.6)	819 (99.9)
Female	687 (69.4)	576 (70.3)
Male	303 (30.6)	243 (29.7)
Age (years)[a]	958 (96.4)	791 (96.5)
18-34	54 (5.6)	50 (6.3)
35-59	269 (28.1)	236 (29.8)
60-74	467 (48.7)	396 (50.1)
≥75	168 (17.5)	109 (13.8)
Education		
Basic education	185 (18.6)	129 (15.7)
Secondary education	523 (52.6)	444 (54.1)
University degree	286 (28.8)	247 (30.1)
Region[a]	992 (99.8)	818 (99.8)
Southern Finland	135 (13.6)	107 (13.1)
Southwestern Finland	144 (14.5)	109 (13.3)
Western and Central Finland	192 (19.4)	155 (18.9)
Eastern Finland	224 (22.6)	189 (23.1)
Northern Finland	222 (22.4)	193 (23.6)
Lapland	75 (7.6)	65 (7.9)
Internet use[a]	987 (99.3)	814 (99.3)
Daily or on several days a week	851 (86.2)	772 (94.8)
Once a week or less often	79 (8)	42 (5.2)
Not at all	57 (5.8)	0 (0)
Internet use to search for health-related information[a]	991 (99.7)	819 (99.9)
Yes	842 (85)	770 (94)
No	149 (15)	49 (6)
Chronic diseases diagnosed by a physician[a]	982 (98.8)	809 (98.7)
Yes	823 (83.8)	682 (84.3)
No	140 (14.3)	113 (14)
Does not know	19 (1.9)	14 (1.7)
Number of regularly used prescription medicines[a]	942 (94.8)	780 (95.1)
0	101 (10.7)	87 (11.2)
1-4	604 (64.1)	496 (63.6)
≥5	237 (25.2)	197 (25.3)
My Kanta use		
Yes	820 (82.5)	820 (100)
Has used it but is not going to use it anymore	21 (2.1)	0 (0)
Has never used it	153 (15.4)	0 (0)

[a]Some of the respondents did not answer the question.

Benefits

Of the My Kanta users, 81.3% (667/820) described at least one benefit that My Kanta use had provided (Table 2). The most frequently stated benefits were opportunities to view one's own health data (290/667, 43.5%) and e-prescriptions (247/667, 37%) and to renew prescriptions (220/667, 33%). Many users (107/667, 16%) also thought it was useful in general that their own information was viewable via the service.

Table 2. Benefits that My Kanta has provided for its users.

Benefits[a]	Respondents (n=667), n (%)
Health data viewable in the service	290 (43.5)
Viewing health data	—[b]
Viewing laboratory results	—
Viewing records of health care visits	—
e-Prescriptions viewable in the service[c]	247 (37)
Monitoring e-prescriptions	—
Monitoring validity of prescriptions	—
Monitoring amounts of outstanding medicines	—
Monitoring the need for prescription renewals	—
Prescriptions are in the service	—
Prescription renewals	220 (33)
Opportunity to renew prescriptions	—
Ease of prescription renewing	—
One's own information viewable in the service	107 (16)
One's own information can be viewed or checked	—
Ease of viewing one's own information	—
One's own information viewable anytime or anywhere	—
One's own information viewable at home	—
All information in one place	—
Fever contacts and calls to health services	42 (6.3)
Fewer calls	—
No need to visit a physician	—
No need to visit a health care center or pharmacy	—
Easy to take care of health-related matters	35 (5.2)
Easy	—
Can be used anytime or anywhere	—
Can be used at home	—
Saving time	—
Other	58 (8.7)
Printing out information	—
Saving organ testament	—
Saving living will	—
Acting on behalf of others	—

[a]From each main category, the most common subcategories are reported in the table.

[b]Not available.

[c]In Finland, all prescriptions are issued electronically.

The main category *Health data viewable in the service* included responses concerning laboratory results, records of health care visits, and health data in general (Table 2). It was thought useful that laboratory results, records of health care visits, and health

data in general were viewable and could be easily monitored via the service. The respondents stated that by using My Kanta, they could keep up to date with their health information. They also found it useful, after their health care visit, to be able to check what had been discussed during the visit or what the physician had recorded.

The most frequent response in the main category *E-prescriptions viewable in the service* was that e-prescriptions could be (easily) viewed and monitored via the service (Table 2). Some users said that they kept a check on how long e-prescriptions were valid, whether there were any medicines outstanding, or whether e-prescriptions needed to be renewed. Some users found it useful that e-prescriptions were included in the service as there was no longer any need to keep paper versions and all prescriptions were safe in one place.

The main category *Prescription renewals* included responses that regarded the opportunity to renew prescriptions via the service as beneficial (Table 2). The renewal process was described as straightforward and fast.

In the main category *One's own information viewable in the service*, the responses focused on the benefits of having access to one's own information in general and that it was useful that information could easily be checked in one place, at anytime and anywhere (Table 2). Many respondents said that it was useful to be able to view this information in the privacy of their homes.

Problems

More than 1 in every 3 users (311/820, 37.9%) described at least one problem in the use of My Kanta (Table 3). The most frequently mentioned problems were that the service lacked health data (71/311, 22.8%) and that users had difficulty navigating within the service and searching for information (68/311, 21.9%). Some users regarded the delay in health data being downloaded into the service to be a problem (41/311, 13.2%).

Table 3. Problems that users have experienced with My Kanta.

Problems[a]	Respondents (n=311), n (%)
The service is lacking health data	71 (22.8)
Laboratory results not available	__[b]
Records of certain health care visits or units not available	—
Old data not available	—
Difficulty navigating the service and searching for information	68 (21.9)
Difficulty finding information	—
The service is badly organized	—
Problems switching between pages	—
Health data viewable with a delay	41 (13.2)
Health data downloaded into the service with a delay	—
Records of health care visits downloaded into the service with a delay	—
Laboratory test results downloaded into the service with a delay	—
Telecommunication problems	33 (10.6)
The service does not open properly	—
The service cuts off the user or is disrupted	—
Problems with the internet connection	—
User-driven challenges in using the service	31 (10)
Illnesses inhibit use of the service	—
Uncertainties about using computer or internet	—
Unaware of the content of the service	—
Difficulty understanding the health data recorded	—
My Kanta and other patient portals get mixed up	—
Inconveniences in logging in	27 (8.7)
Logging in is not possible without web banking codes	—
Web banking codes are not always carried	—
Web banking does not work properly	—
Logging in with web banking codes is laborious	—
Difficulty in prescription monitoring	23 (7.4)
Invalid and valid prescriptions become mixed	—
Difficulty understanding valid dates of prescriptions or amounts of outstanding medicines	—
Prescriptions cannot be arranged	—
Inconveniences in renewing prescriptions	21 (6.8)
Certain[c] prescriptions cannot be renewed	—
Physician has not renewed the prescription	—
Problems choosing a health care unit for renewal	—
Incorrect information in the service	13 (4.2)
Information recorded incorrectly	—
Other persons' data in the service	—
Difficulty correcting erroneous information	—
Other	37 (11.9)
Difficulty printing out e-prescriptions and health data	—

Problems[a]	Respondents (n=311), n (%)
Guardian cannot see the data of dependents who are younger than 10 years	—
Uncertainties in data sharing procedures	—

[a]From each main category, the most common subcategories reported in the table.

[b]Not available.

[c]For example, prescriptions issued in private health care or prescriptions issued over 28 months ago.

The main category *The service is lacking health data* included responses that My Kanta is lacking certain data (Table 3). Users reported that laboratory results were either partly or completely unavailable. In addition, some users reported that the records of certain health care visits or units did not seem to have been entered into the system. Some responses said that records of private health care providers, such as occupational health care, were missing. Users also reported that some information had disappeared, or they would like to see their health information from the years before the introduction of My Kanta.

Difficulty finding information in the service was frequently reported to be a problem in the main category *Difficulty navigating the service and searching for information* (Table 3). Some respondents did not specify what information they could not find, whereas others mentioned laboratory results or records of a particular health care visit. Some users thought that the service was badly arranged. For example, layout, menus, and headings were considered unclear, which complicated their ability to navigate within the service. Some users reported that switching between different pages did not work properly, as they could not move back to the previous page but instead were transported to the starting page.

Responses in the main category *Health data viewable with a delay* refer to the feeling that there are delays in health data being downloaded into the service (Table 3). Some respondents did not specify what health data they meant, whereas others mentioned records of health care visits or laboratory results.

Potential Improvements

Of My Kanta users, 39.9% (327/820) reported at least one potential improvement in the service (Table 4). The most common suggestions for improvements were that the health data provided in the portal should be more comprehensive (89/327, 27.2%) and navigating the service and searching for information should be made easier (71/327, 21.7%). Some users suggested new functions that could be incorporated into the service (51/327, 15.6%) or would like health data to be entered more promptly into the service (47/327, 14.4%).

Table 4. Potential improvements proposed by users.

Proposed improvements[a]	Respondents (n=327), n (%)
More comprehensive health data	89 (27.2)
Data from all health care units should be available	—[b]
All data viewable in the service	—
Records of health care visits in more detail	—
Old data should be incorporated into the service	—
Facilitating navigation in the service and searching for information	71 (21.7)
Data grouping more clearly	—
Clearer layout	—
Simplification of the service	—
Use guidance	—
More functions incorporated into the service	51 (15.6)
Secure messaging	—
Notifications	—
Vaccinations	—
More information about medicines	—
Opportunity to correct erroneous information	—
Faster downloading of health data	47 (14.4)
Laboratory results entered into the service with a shorter delay	—
Record of health care visits into the service with a shorter delay	—
Data entered into the service with a shorter delay	—
Easier monitoring of prescriptions	33 (10.1)
Certain[c] prescriptions should be removed	—
Clearer order for prescriptions	—
Clearer prescription information	—
Health data in plain language	23 (7)
Diagnosis, terms, and medical reports in easy-to-comprehend language	—
Laboratory results in easy-to-comprehend language	—
Easier logging in and mobile use	22 (6.7)
Other ways[d] to log-in than web banking codes	—
Simplifying the log-in procedure	—
Mobile app should be introduced	—
Making it easier to act on behalf of others	19 (5.8)
Opportunity to act on behalf of dependents aged 10-18 years[e]	—
Opportunity to act on behalf of adults (>18 years)[e]	—
Other	26 (8)
Improvements in prescription renewals	—
Improvements in printing out information	—
Data protection	—

[a]From each main category, the most common subcategories are reported in the table.

[b]Not available.

[c]For example, invalid or noncurrent prescriptions.

[d]For example, username and password or PIN code.

[e]In Finland, the age of majority is 18 years.

Responses in the main category *More comprehensive health data* related to health data that is or should be in the portal (Table 4). These responses included the view that health data from all visits or units should be added to the service. Some respondents did not specify from which units' data were missing, whereas others mentioned as an example of private health care providers or oral health care. Many users wanted the laboratory results to be viewable in the service. Some of these respondents had no laboratory results in the service, whereas for others, they were only partially available. Some users wanted records of health care visits to be more detailed. A few respondents wanted data from the years before My Kanta or data that had disappeared to be added to the service.

In the main category *Facilitating navigation in the service and searching for information*, some respondents wanted the data to be grouped more clearly (Table 4). It was also suggested that there could be clear sections for different data, such as laboratory results, data from certain units, or the latest data. Some respondents wanted data to be arranged in a clearer chronological order. Some thought that searching for particular information should be made easier and the layout of the service clearer. Specifically, the headings and menus were considered to be unclear. Some users wanted the service to be simpler. Others reported that they would like to have more guidance in using the service. A few would like to receive face-to-face guidance, whereas others thought that some kind of manual or more general information should be provided.

The main category *More functions incorporated into the service* included responses proposing the addition of new functions to the portal (Table 4). Some users wanted secure messaging between patients and health care providers. It was considered an advantage if patients could ask physicians or nurses about their care or medicines via the service, or, in the case of prescription renewals, if they could send messages to their physician. Respondents also proposed the introduction of novel notifications, for example, an SMS text message to a mobile phone informing them when prescriptions had to be renewed or when new records were downloaded to the portal. Some users wanted more information about medicines to be included in the portal, such as drug interactions or costs. It was also suggested that vaccinations should be included. Some users wanted to have an opportunity to correct erroneous information via the service or to report incorrect information.

In the main category *Faster downloading of health data*, some respondents wanted to see laboratory results, records of health care visits, or data in general to be viewable in the service more promptly (Table 4). A few respondents wanted data to be entered into the service without delay or with a shorter delay, so they could use the information at their next appointment.

Discussion

Principal Findings and Comparison With Prior Work

In this study, more pharmacy customers reported benefits than problems or potential improvements when using the nationwide patient portal My Kanta. Users consider My Kanta beneficial because health and prescription information can be easily viewed and monitored, and prescription renewals can be submitted via the service. In general, the fact that all these services are available regardless of time and place, for example, in the privacy of the patient's home, is regarded as a benefit. However, the system is not perfect, and some drawbacks were reported. Ideas for improvement mainly concerned certain frequently mentioned problems. The results of this study are mostly in line with those of previous studies investigating users' experiences regarding the benefits, problems, and potential improvements of patient portals [3-11]. This suggests that there are no major differences in users' experiences between nationwide and organization-specific portals, although these portals differ slightly from each other in terms of content and functions.

The My Kanta functions rated most beneficial, that is, viewing health data and prescriptions and renewing prescriptions, are also the most widely used [16]. According to a study conducted in 2014 among the Finnish population on the use of electronic services in health care, respondents said that in the future, they would like to have access to their laboratory test results, health records, and prescriptions, and an opportunity to submit renewal requests [24]. Thus, as these options are now available, individuals naturally regard them as the most beneficial, and these functions are most extensively used. The present results also indicate that My Kanta met users' expectations well. In addition, the portal has achieved its intended goal of making it easier for patients to obtain their health information and hence participate in their own care.

In line with previous studies [3-7,9-11], the ability to view one's own health data was regarded as the most beneficial function of the patient portal. However, users reported some problems that may discourage the use of health data and thus patient engagement. Some users found the language used in their health records too difficult to understand. This is also a finding from previous studies conducted in Finland [16] and other countries [3,5,10,11]. My Kanta is a nationwide portal that can be used by anyone living in Finland with an ID for electronic services. It is important that the language used in the portal is plain enough to allow everyone to easily understand the recorded data. Physicians should be informed about this finding so that they can avoid writing reports with professional terminology that laypeople will not understand. It would also be useful if My Kanta contained explanations of laboratory test abbreviations and reference values.

To increase and facilitate patient engagement, there are other issues that require attention. Many users reported that not all their health data, especially laboratory results and data from certain health care units, were available. This prevents patients from gaining a comprehensive picture of their health and care. Health data from public health care started to be included in My Kanta in 2013 and from private health care in 2016 [12]. It may be that at the time of this study, not all units had started to record their data in My Kanta, as this took place step by step. It should also be noted that the portal is continuously developing

and that the amount of health data that can be accessed is constantly increasing [26]. For example, since the questionnaires were returned, oral health care data and vaccinations have begun to be downloaded to the portal. Furthermore, there is a possibility that data on some health care visits or units are delayed as health care professionals want to discuss the significance of findings with the patient before the data are made available in My Kanta [27]. In Finland, there is no exact time period determining when certain data should be entered in My Kanta; instead, this is left to the discretion of health care units or professionals. In some cases, records can be delayed indefinitely if their availability is seen as a risk to the patient's life or care. It would be important to inform patients about these delay procedures as many users wanted their health data to be downloaded to My Kanta more promptly. The wish of patients to have their health data downloaded to the portal without delay has also emerged in studies conducted in other countries [3,9,11]. In contrast, in a study conducted in the Netherlands, users did not want their health data to be downloaded to the portal before they had visited a physician [5]. Some users may not have the ability to understand information, such as the meaning of laboratory results, and thus, delaying the data may be necessary to prevent unnecessary concerns. Nevertheless, it may be that patients are unaware of the reasons for a delay, and they should be informed accordingly, for instance, when visiting the health care provider.

Users also stated that some data were difficult to find in My Kanta. This was also highlighted in a previous study that investigated My Kanta usability [16]. In this study, many suggestions were made to make searching for information easier, for example, organizing data in the portal in different ways. Users also wanted the service to be simpler, especially headings. However, many of the suggestions differed from each other, so the first step in making information easier to find could be better instructions on the use of the portal, which was also the wish of users. It is evident that further investigations should be conducted to determine the optimal information search strategy. At the moment, Kanta pages on the internet include a considerable amount of self-guidance. There is a My Kanta web-based course including guidance videos and instructions to acquaint users with My Kanta functions and how to start using the portal [28]. Clients can also request help from the customer service [29]. However, as the wishes for easier use of the portal and more guidance were mentioned in many responses and as some users reported they were unaware of the content of My Kanta, it can be assumed that this information and guidance have not reached all citizens. A wish for further guidance on My Kanta use was also reported in a previous study [16]. As suggested in this study, it may be necessary to provide face-to-face guidance or a manual.

Opportunities to view and renew prescriptions emerged as the most commonly experienced benefits, which is a finding at odds with previous studies [3,5,7,10]. This difference may be due to the fact that there are only few countries where the use of e-prescriptions is as widespread as in Finland, where all prescriptions have been issued electronically since 2017 [30]. In Finland, My Kanta is the only way for patients to view and monitor their prescriptions, other than by visiting or calling health care units or pharmacies. In addition, the survey respondents were pharmacy customers who used prescription medicines, which may have biased the result. Although users experienced the ability to view prescriptions as beneficial, some suggested ways to make prescription monitoring easier. Respondents wanted to see prescriptions not in use removed from the service or the opportunity to arrange prescriptions or hide unnecessary prescriptions. This would simplify the overall view of their prescriptions in the service. Kanta Services is developing a national medication list [31]. This will be an up-to-date list of currently used medicines to be included in My Kanta, which may help patients to obtain an overview of their medication. Users also had difficulty understanding prescription information, especially valid dates and amounts of outstanding medicines. Health care professionals should explain this information to patients in health care units and pharmacies.

In this study, prescription renewal via My Kanta was experienced as easy, and some users stated that they now visited or called health care providers less often because they were able to send renewal requests via the portal. This saves time for health care professionals, which is one of the aims of eHealth services [2]. However, the renewal of prescriptions without any communication between the patient and health care professionals may pose risks for pharmacotherapy monitoring. More studies are therefore needed on prescription renewal via My Kanta so that the overall impact of this function on health care professionals' workload and pharmacotherapy monitoring can be evaluated.

Users also wanted to see new information or functions included in the portal. This is in line with previous studies [3,4,8,10,11], indicating that portal users want all their health-related data to be available via these services. It could be argued that users would prefer patient portals to be web-based places where they could deal with many aspects of their health and medication. Communicating with health care professionals and notifications were the most often desired new functions. In Finland, some private health care providers have their own organization-based patient portals, which include communication features. It is common worldwide for these functions to be part of organization- or area-specific portals, and when available, these are also some of the most beneficial functions reported by portal users [3-7,11]. However, they have not been implemented in nationwide portals [9,10]. A communication function could be a useful way to enhance patient engagement and decrease the need for contact with health care. However, this might increase the workload of health care professionals, and therefore, before such a function is implemented in a nationwide patient portal, its impacts would need to be scrutinized in detail. The possibility of including notifications in My Kanta, as suggested in this survey, should also be evaluated. Users also wanted an opportunity to correct erroneous information on the portal. Some users reported that they had noted that their own data were incorrect or that there was some other person's data in their portal. This problem was also mentioned in a previous study on My Kanta use [16]. In Finland, the procedure for correcting erroneous information is for the patient to contact the health care unit where the incorrect information has been entered and

ask them to correct the error [13]. Evidently, this procedure is regarded as inconvenient.

Strengths and Limitations

This study has both strengths and limitations. The findings are based on a nationwide patient portal that has been widely adopted by the Finnish population. By handing out the questionnaires from the pharmacies after dispensing the prescription medicine, we reached our target group (ie, medicine users who potentially have a need to use My Kanta). People who have prescription medicines also have contacts in health care and hence health data recorded in the service. Thus, medicine users represent a population with both health data and prescriptions recorded in My Kanta. This makes it a good population to study the use of My Kanta service. However, this target group may bias the results concerning prescriptions, as all respondents used prescription medicines.

The study sample was large and included pharmacy customers from all parts of the country. Evaluating the representativeness of the results is challenging because there are no comparable statistics on the characteristics of Finnish pharmacy customers. As the customers were recruited anonymously, we have no knowledge of the characteristics of those who declined to participate or those who did not return the questionnaires. The response rate in this study (994/2866, 34.68%) was low. However, the respondents' characteristics (age, gender, education, and region) were fairly similar to those in earlier Finnish studies of pharmacy customers conducted using the same method with better response rates (40%-44%) [18,32].

The questions reported in this study were not validated measures. However, the questions were based on previous studies, with minor revisions [25,33]. In addition, both face validity and pilot tests were conducted. Many users (311/820, 37.9%-667/820, 81.3%) responded to the open-ended questions presented in this study. The reliability of categorization was ensured by the fact that 2 researchers conducted an inductive content analysis. Contradictory categorizations were discussed in the research group to obtain a consensus.

Conclusions

Finnish pharmacy customers experience more benefits than problems or potential improvements when using the nationwide patient portal My Kanta. They perceive the portal as an easy and beneficial way to view their health data and prescriptions and to renew their prescriptions. However, portal users wanted the service to include more information and functions, and for data searches to be made easier. Fulfilling these wishes could encourage even greater use of the portal and help patients use the data recorded there. These improvements could allow patients to become more involved in their own health and care.

Acknowledgments

The authors would like to thank the pharmacies for their cooperation in distributing the questionnaires. This study was funded by the Social Insurance Institution of Finland (Kela). The statements presented in this document are those of the authors and do not reflect the official position of Kela.

Authors' Contributions

MS, RA, and JT designed the study and collected the data. MS and AO conducted the analyses and drafted the first version of the manuscript. MS drafted the final version of the manuscript. All authors participated in discussing the analyses and findings, critically revised the manuscript, and read and approved the final version to be submitted. AO was affiliated with the University of Eastern Finland at the time of the study and is currently affiliated with Humalisto Pharmacy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire survey for pharmacy customers regarding the My Kanta service.
[DOC File , 143 KB - jmir_v23i11e31483_app1.doc]

References

1. World Health Organization. Patient Engagement: Technical Series on Safer Primary Care. Geneva: World Health Organization; 2016.
2. World Health Organization. Atlas of eHealth Country Profiles: the Use of Ehealth in Support of Universal Health Coverage: Based on the Findings of the Third Global Survey on eHealth, 2015). Geneva: World Health Organization; 2016.
3. Wildenbos GA, Maasri K, Jaspers M, Peute L. Older adults using a patient portal: registration and experiences, one year after implementation. Digit Health 2018;4:2055207618797883 [FREE Full text] [doi: 10.1177/2055207618797883] [Medline: 30186619]
4. Son H, Nahm E. Older adults' experience using patient portals in communities: challenges and opportunities. Comput Inform Nurs 2019 Jan;37(1):4-10. [doi: 10.1097/CIN.0000000000000476] [Medline: 30188353]

5. Vreugdenhil MM, Ranke S, de Man Y, Haan MM, Kool RB. Patient and health care provider experiences with a recently introduced patient portal in an academic hospital in the Netherlands: mixed methods study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
6. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
7. Ronda MC, Dijkhorst-Oei L, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
8. Wass S, Vimarlund V. The role of PAEHRs in patient involvement. J Med Syst 2018 Sep 25;42(11):210 [FREE Full text] [doi: 10.1007/s10916-018-1070-y] [Medline: 30255404]
9. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
10. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
11. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
12. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. FinJeHeW 2018 Dec 04;10(4):381-395. [doi: 10.23996/fjhw.74511]
13. My Kanta Pages. Kanta Services. URL: https://www.kanta.fi/en/my-kanta-pages [accessed 2021-04-07]
14. Acting on behalf of a minor. Kanta Services. URL: https://www.kanta.fi/en/acting-on-behalf-of-a-minor [accessed 2021-04-07]
15. Buying medications abroad. Kanta Services. URL: https://www.kanta.fi/en/web/guest/buying-prescription-medicines-abroad [accessed 2021-04-07]
16. Sääskilahti M, Ahonen R, Timonen J. Pharmacy customers' experiences of use, usability, and satisfaction of a nationwide patient portal: survey study. J Med Internet Res 2021 Jul 6;23(7):e25368 [FREE Full text] [doi: 10.2196/25368] [Medline: 34269687]
17. Kanta services popular among citizens and professionals – year 2019 in a nutshell. Kanta Services. URL: https://www.kanta.fi/en/notice/-/asset_publisher/cf6QCnduV1x6/content/kanta-palvelut-kansan-ja-ammattilaisten-suosiossa-talta-naytti-vuosi-2019 [accessed 2021-05-10]
18. Lämsä E, Timonen J, Mäntyselkä P, Ahonen R. Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland. Int J Med Inform 2017 Jan;97:221-228. [doi: 10.1016/j.ijmedinf.2016.10.014] [Medline: 27919380]
19. Sääskilahti M, Aarnio E, Lämsä E, Ahonen R, Timonen J. Use and non-use of a nationwide patient portal – a survey among pharmacy customers. J Pharm Health Serv Res 2020 Jul 12;11(4):335-342. [doi: 10.1111/jphs.12368]
20. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663 [FREE Full text] [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
21. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
22. van der Vaart R, Drossaert CH, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MA. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
23. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
24. Hyppönen H, Hyry J, Valta K, Ahlgren S. Electronic services in the social welfare and health care sector. Citizens' experiences and development needs. Report 33/2014. Finnish Institute for Health and Welfare (THL). 2014. URL: http://www.julkari.fi/handle/10024/125597 [accessed 2021-06-16]
25. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 03;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
26. Updates in My Kanta Pages. Kanta Services. URL: https://www.kanta.fi/en/updates-in-my-kanta-pages [accessed 2021-04-07]
27. Frequently asked questions. Kanta Services. URL: https://www.kanta.fi/en/faq [accessed 2021-06-16]
28. Get to know your My Kanta Pages. Kanta Services. URL: https://www.kanta.fi/en/my-kanta-pages-online-school [accessed 2021-06-04]
29. Support. Kanta Services. URL: https://www.kanta.fi/en/support [accessed 2021-06-04]

30. The act on electronic prescription 61/2007. Finlex. 2007. URL: https://www.finlex.fi/fi/laki/ajantasa/2007/20070061 [accessed 2021-06-16]
31. Priorities of the Kanta Services. Kanta Services. URL: https://www.kanta.fi/en/web/guest/professionals/priorities-of-the-kanta-services [accessed 2021-06-10]
32. Nokelainen H, Lämsä E, Ahonen R, Timonen J. Reasons for allowing and refusing generic substitution and factors determining the choice of an interchangeable prescription medicine: a survey among pharmacy customers in Finland. BMC Health Serv Res 2020 Feb 03;20(1):82 [FREE Full text] [doi: 10.1186/s12913-020-4894-3] [Medline: 32013951]
33. Sääskilahti M, Ahonen R, Lämsä E, Timonen J. Sähköisen reseptin edut ja ongelmat - kyselytutkimus apteekkien asiakkaille. [Benefits and problems of electronic prescriptions – a survey among pharmacy customers in Finland]. DOSIS 2016;32(2):129-141 [FREE Full text]

Edited by R Kukafka; submitted 30.06.21; peer-reviewed by R Marshall, S MacEwan, M Avdagovska; comments to author 04.08.21; revised version received 31.08.21; accepted 12.09.21; published 03.11.21.

Please cite as:
Sääskilahti M, Ojanen A, Ahonen R, Timonen J
Benefits, Problems, and Potential Improvements in a Nationwide Patient Portal: Cross-sectional Survey of Pharmacy Customers' Experiences
J Med Internet Res 2021;23(11):e31483
URL: https://www.jmir.org/2021/11/e31483
doi:10.2196/31483
PMID:34730542

Original Paper

Patients Contributing to Visit Notes: Mixed Methods Evaluation of OurNotes

Jan Walker[1,2], MBA, RN; Suzanne Leveille[1,3], PhD, RN; Gila Kriegel[1], MD; Chen-Tan Lin[4], MD; Stephen K Liu[5], MPH, MD; Thomas H Payne[6], MD; Kendall Harcourt[1], MPA; Zhiyong Dong[1], MSc; Patricia Fitzgerald[1], MSc; Matthew Germak[7], MPH, MD; Lawrence Markson[8], MPH, MD; Sara L Jackson[6], MPH, MD; Hannah Shucard[9], MSc; Joann G Elmore[10], MPH, MD; Tom Delbanco[1], MD

[1]Division of General Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]College of Nursing and Health Sciences, University of Massachusetts Boston, Boston, MA, United States
[4]School of Medicine, University of Colorado, Aurora, MA, United States
[5]General Internal Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, MA, United States
[6]Department of Medicine, University of Washington School of Medicine, Seattle, MA, United States
[7]Primary Care, Beth Israel Lahey Health, Needham, MA, United States
[8]Clinical Information Systems, Beth Israel Deaconess Medical Center, Boston, MA, United States
[9]Department of Biostatistics, University of Washington School of Medicine, Seattle, MA, United States
[10]David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, MA, United States

Corresponding Author:
Jan Walker, MBA, RN
Division of General Medicine
Beth Israel Deaconess Medical Center
331 Brookline Avenue
Boston, MA, 02215
United States
Phone: 1 6173208733
Email: jwalker1@bidmc.harvard.edu

Abstract

Background: Secure patient portals are widely available, and patients use them to view their electronic health records, including their clinical notes. We conducted experiments asking them to cogenerate notes with their clinicians, an intervention called *OurNotes*.

Objective: This study aims to assess patient and provider experiences and attitudes after 12 months of a pilot intervention.

Methods: Before scheduled primary care visits, patients were asked to submit a word-constrained, unstructured interval history and an agenda for what they would like to discuss at the visit. Using site-specific methods, their providers were invited to incorporate the submissions into notes documenting the visits. Sites served urban, suburban, and rural patients in primary care practices in 4 academic health centers in Boston (Massachusetts), Lebanon (New Hampshire), Denver (Colorado), and Seattle (Washington). Each practice offered electronic access to visit notes (open notes) to its patients for several years. A mixed methods evaluation used tracking data and electronic survey responses from patients and clinicians. Participants were 174 providers and 1962 patients who submitted at least 1 previsit form. We asked providers about the usefulness of the submissions, effects on workflow, and ideas for the future. We asked patients about difficulties and benefits of providing the requested information and ideas for future improvements.

Results: Forms were submitted before 9.15% (5365/58,652) eligible visits, and 43.7% (76/174) providers and 26.76% (525/1962) patients responded to the postintervention evaluation surveys; 74 providers and 321 patients remembered receiving and completing the forms and answered the survey questions. Most clinicians thought interim patient histories (69/74, 93%) and patient agendas (72/74, 97%) as good ideas, 70% (52/74) usually or always incorporated them into visit notes, 54% (40/74) reported no change in visit length, and 35% (26/74) thought they saved time. Their most common suggestions related to improving notifications when patient forms were received, making it easier to find the form and insert it into the note, and educating patients about how best to prepare their submissions. Patient respondents were generally well educated, most found the history (259/321, 80.7%)

and agenda (286/321, 89.1%) questions not difficult to answer; more than 92.2% (296/321) thought sending answers before the visit a good idea; 68.8% (221/321) thought the questions helped them prepare for the visit. Common suggestions by patients included learning to write better answers and wanting to know that their submissions were read by their clinicians. At the end of the pilot, all participating providers chose to continue the *OurNotes* previsit form, and sites considered expanding the intervention to more clinicians and adapting it for telemedicine visits.

Conclusions: *OurNotes* interests patients, and providers experience it as a positive intervention. Participation by patients, care partners, clinicians, and electronic health record experts will facilitate further development.

(J Med Internet Res 2021;23(11):e29951) doi:10.2196/29951

KEYWORDS

electronic health record; previsit information; physician-patient relations; patient portal; mobile phone

Introduction

Background

Patient engagement is essential for improving patient experience, reducing the cost of care, and improving population health [1]. With the proliferation of secure patient portals over the past decade, patients have gained a powerful engagement tool [2,3]. Millions of patients and their care partners now go on the web to communicate with members of the care team and review their medical records. Beginning in April 2021, as part of the 21st Century Cures Act, patients nationwide have access to virtually all information in their electronic health records, including the notes of their clinicians [4].

Hundreds of provider organizations implemented *open notes* on their patient portals following publication of the first open notes study in 2012 [5,6]. Patients rate note reading as very important for remembering what happened in their office visits, remembering their care plan, taking care of their health, and feeling in control of their care [7-10]. Patients have also expressed interest in doing more than passively reading their notes. In the first study of open notes, 60% wanted to be able to comment on notes, and 35% wanted to be able to approve their notes [5]. Though few primary care physicians thought patients should approve the content of notes, approximately one-third agreed they should be able to comment. Researchers have since shown that reading notes and inviting patients to comment on their notes can benefit patient safety [11-16]. Though many clinicians face time pressures, a majority report spending the same or less time writing open notes [17]. The National Academy of Medicine's Vital Directions for Health and Health Care initiative recommends that patients and authorized family caregivers be able to modify their health records [18]. Though many ambulatory care practices collect information from patients in brief questionnaires, we know of no efforts inviting them to actively write in their records. As a step in this direction, we decided to mount a pilot inviting patients to contribute to their visit notes, thus transforming the official record of their care into cogenerated *OurNotes*.

We designed, implemented, and evaluated portal mechanisms for patients and their care partners to contribute to notes of providers documenting their ambulatory visits. Following advice from expert interviews, we determined to focus on previsit communications by asking patients to (1) write a brief history since the last visit and (2) to identify up to three priorities for the upcoming visit [19].

Objectives

Our objectives were to have providers incorporate this previsit input into their visit notes, thereby taking a first step toward *OurNotes*, and to assess patient and provider experiences with this pilot intervention. We hypothesized that clinicians and patients would find it beneficial to submit the information before visits, especially patients with a heavy burden of illness, and that the intervention would be time-neutral or possibly save time for clinicians. This paper reports the principal findings from 12-month pilots of this intervention in 4 sites.

Methods

Overview

Primary care practices associated with 4 sites across the United States participated: Beth Israel Deaconess Medical Center (BIDMC) in Boston, UCHealth (UCH) in Aurora, Colorado (including a women's health clinic), Dartmouth-Hitchcock Medical Center (DHMC) in Lebanon, New Hampshire, and University of Washington Medicine (UW) in Seattle (including a safety net clinic and a clinic that serves primarily privately insured patients). Each site has well-established mechanisms for offering open notes to patients. The *OurNotes* and site project teams developed an overall intervention design and evaluation plan. The study protocols were approved by the institutional review boards of the participating institutions.

Intervention

The intervention included primary care providers and patients who were registered on the patient portal. During the 12-month pilot period, these patients were invited via a secure portal message or email to complete a form addressing two questions before a scheduled visit: *How have you been since your last visit?What are the most important things you would like to discuss at your visit? (list up to 3)*. Answers were limited to 2000 and 300 characters, respectively. Clinicians could view the completed forms before and during the visit, and when they wrote the visit note, they could decide whether and to what degree to incorporate the patient's words.

Between mid-2017 and early 2019, each site built its own implementation process and supported the information technology infrastructure. Although all asked the same two questions up to 1 week before the visit, the sites differed in their

approaches (Multimedia Appendix 1). Notably, 2 sites implemented practice-wide approaches with many clinicians (UC and UW), and 2 recruited volunteer clinicians (BIDMC and DHMC). Two sites (UC and UW) used Epic (Epic Systems Corporation, Verona, Wisconsin) to facilitate the insertion of patients' submissions into notes. They were able to launch 9 months earlier than the other 2 sites.

In all sites, invitations to patients (Multimedia Appendix 2) included links to the form (Multimedia Appendix 3). The invitation described the intervention, explained that completing the form was voluntary, and noted that the form might not be read if they did not keep the appointment. At BIDMC, UC, and UW, automated processes sent the previsit invitations; physicians at DHMC selected patients from the weekly schedule, and clinic staff or the physician sent the invitation message.

Evaluation

Pilot interventions were conducted between June 2018 and April 2020. Three sites (BIDMC, UC, and UW) used portal tracking data to monitor the use of the form throughout the pilot, and clinicians at DHMC made periodic estimates. The evaluation plan also included surveys of both clinicians and patients at each site and moderated discussions with the selected groups. However, the COVID-19 pandemic led to restrictions on research activities and diversion of information systems resources to support clinical care. Therefore, we were unable to complete all parts of the planned evaluation. We completed both patient and clinician surveys at the 2 sites that finished their 12-month pilots in 2019 (UC and UW), and surveyed clinicians at the other 2 sites (BIDMC and DHMC) in Spring 2020. However, without information systems resources to identify and track participants in early 2020, we were unable to conduct patient surveys at BIDMC or DHMC. We were also unable to convene focus groups.

The research team developed survey items based on pilot objectives, informal comments, and feedback from each site. We asked clinicians about the usefulness of patient responses, effects on workflow, and ideas for the future. We asked patients about the difficulties and benefits of answering the questions and about ideas for future improvements. To assess the burden of illness, we asked the patients if they had a chronic illness. The surveys included brief sets of sociodemographic items and both closed-ended and free-text questions. The questionnaires are available on request.

We invited clinicians who had received at least three completed previsit forms in the previous 6 months to complete the survey. We invited patients at UC and UW who had submitted at least one form during the pilot study. Clinicians in the 4 sites were contacted between June 2019 and July 2020, and UC and UW patients were contacted between September and October 2019. All participants received an invitation and up to two reminders. Each site offered an incentive of US $10 to US $50, following local guidelines. The surveys were conducted using Research Electronic Data Capture (REDCap, Vanderbilt).

Analysis

Descriptive statistics were used to summarize response rates for the survey, survey responses, and the proportion of visits for which patients returned forms (using portal tracking data at BIDMC, UC, and UW, and clinician estimates at DHMC). In addition, patient results were stratified according to whether they reported having a chronic illness.

We completed thematic analyses of the free-text comments from the surveys. Patients were asked two questions: one for ways to improve the *OurNotes* previsit process, and one for other comments. To analyze responses, 2 authors first scanned the responses and noted that both questions drew comments related to both *OurNotes* and to other topics; therefore, we combined the comments for analysis. We reviewed 30 comments independently and drafted six topics, then independently reviewed 30 new comments and defined the final set of themes for coding: comments about the two questions, the form, workflow, patient engagement, whether the submitted form was read by clinicians, and others. Using REDCap, both authors coded all responses with up to three codes each, and 1 selected representative examples.

Clinicians were asked a single question about how to improve the previsit *OurNotes* process and any other comments. Two authors reviewed 30 comments to establish a list of nine themes, then 1 author reviewed all comments, assigned 1 or more themes to each, and selected representative thematic examples.

Results

Overview

In the 3 sites with portal tracking, patients were invited to complete forms before 58,652 scheduled visits, 5365 (9.1%) forms were submitted, and in the site without portal tracking, forms were returned before approximately 50% (260/520) visits (Multimedia Appendix 1). We invited 174 clinicians to complete the evaluation survey, and 76 (43.7%) responded. Two respondents did not remember receiving any completed forms, leaving an analytic data set of 74 clinicians. We contacted 1962 patients who had submitted at least one form and 525 (26.8%) responded; 204 did not remember completing a form, leaving an analytic data set of 321 patients.

Clinician Results

Among the 74 clinician respondents, 49 (66%) were women, 61% (45/74) obtained their licenses since 2000, and 54% (40/74) reported 21 or more patient visits per week. On average, clinicians reported receiving between 1 and 3 patient forms per week in the previous 3 months (Table 1). Nearly all respondents agreed that having the information from the forms was a good idea, and 70% (52/74) reported that they usually or always incorporated responses of patients in the visit notes. Almost 90% (66/74) of clinicians reported that having the previsit information of patients either saved time or did not change the visit length.

Table 1. Responses from 74 clinicians about their experiences with OurNotes.[a]

Survey question	Response options	Value, n (%)
In the last 3 months, about how many questionnaires did you receive from patients?		
	<1/week	17 (23)
	1-3/week	38 (51)
	4-10/week	19 (26)
	>10/week	0 (0)
In general, receiving the interim history before a visit is a good idea.		
	Agree or somewhat agree	69 (93)
	Disagree or somewhat disagree	5 (7)
In general, receiving the patient agenda before a visit is a good idea.		
	Agree or somewhat agree	72 (97)
	Disagree or somewhat disagree	2 (3)
How often did you incorporate some or all of what the patient wrote into your visit note?		
	Never or rarely	12 (16)
	Sometimes	10 (14)
	Usually or always	52 (70)
How having patients' answers to the questions affected time spent writing notes?		
	Took longer	4 (5)
	No change	61 (82)
	Took less time	9 (12)
How useful were interim histories sent by patients?		
	Not useful	7 (9)
	Somewhat useful	50 (68)
	Very useful	17 (23)
How useful to have patients' visit priorities at start of visit?		
	Not useful	1 (1)
	Somewhat useful	31 (42)
	Very useful	38 (51)
	Not available at start	4 (5)
Recall one or more instances where a patient's submission was very important?		
	Yes	25 (34)
	No	49 (66)
Did it change the time to complete a patient encounter?		
	Increased time	8 (11)
	No change	40 (54)
	Saved time	26 (35)

[a]Table excludes 2 respondents who did not remember receiving previsit forms.

The process of entering the responses of a patient to the record varied among the participating sites. In 2 sites where the previsit information could be automatically inserted into the visit note, 92% (45/48) preferred this approach. At the other 2 sites, most clinicians copied and pasted the information, 48% (11/23), read or typed all or part of the words of patients into the note, 13% (3/23), or simply referred to having read what the patient wrote, 30% (7/23).

Clinicians were asked to describe their experiences with the intervention and to suggest changes or improvements. Their 57 text responses centered on better notifications about completed forms, the need for easy insertion of patient text into the note, and ways to support patient participation (Textbox 1). Clinicians

suggested that patients could be encouraged to be more engaged and taught about how to complete the form so that it would be more helpful for the visit.

Textbox 1. Examples from 57 clinicians' responses to: "What changes/improvements would you suggest for the process tested in this pilot? Did it help or hinder you in providing care? Do you have other comments?"

Timely notifications to clinicians

- Send the questionnaire results to my in-basket so I am more likely to see them before the visit.
- There were times I only learned of the patient's written comments from the patient during the visit. I think the patient was disappointed that I had not seen the comments in advance.

Incorporating patient's writing into note

- It would be great if there was an easier way to incorporate patient notes into my own note.
- Improve the way the content of the message is incorporated into the body of the note: it would ideally blast in inside of quotation marks at whatever point the cursor is in the note body.

Supporting patient participation

- This could be very useful but for it to pay off would require "training " my patients so that the responses are helpful to both patient and provider.
- Having this done on a regular basis would require a culture change where patients are less passive and more engaged and proactive about their care and ultimately get more out of each clinic visit.
- I do hope patients receive some type of disclaimer that their doctor might not be able to address all their concerns in a visit, even if the patient lists them all.
- Integration with our existing previsit questionnaires.

Other

- Good system. Please keep it very simple for patients and providers.
- [Customize for] differences between annual visits, [follow-up] visits, and urgent visits.

Patient Results

Among the 321 patients who completed the survey from 2 sites (UC and UW), 57.9% (186/321) were women, 79.7% (255/320) were White, and 73.1% (234/320) had completed a 4-year college degree (Table 2). Half (158/320, 49.3%) of the respondents reported having three or more visits to their primary care provider in the previous 12 months, 68.2% (219/321) reported having one or more major chronic conditions, and 77.2% (247/320) reported their health as excellent, very good, or good. There were very few differences between respondents at the 2 sites (data not shown), other than more women responding in Colorado than in Washington (74% and 47% of respondents, respectively), and a higher proportion of respondents reporting White race in Colorado versus Washington (88% vs 74%, respectively).

Table 2. Self-reported characteristics of 321 patient respondents.[a]

Patient characteristic	Values, n (%)
Age (years)	
18-44	78 (24.4)
45-64	102 (31.9)
≥65	140 (43.7)
Sex	
Female	186 (58.1)
Male	123 (38.4)
Transgender	5 (1.6)
Prefer not say	6 (1.9)
Race	
White	255 (79.7)
Black	11 (3.4)
Asian, Hawaiian, or Pacific Islander	25 (7.8)
Other	13 (4.1)
Multiple races	16 (5.0)
Ethnicity	
Non-Hispanic	300 (94.0)
Hispanic	19 (6.0)
Education	
Masters or doctoral degree	126 (39.4)
4-year degree or some grad school	108 (33.7)
Some college or technical school	72 (22.5)
High school or less	14 (4.4)
Overall health	
Excellent, very good, or good	247 (77.2)
Fair or poor	73 (22.8)
Chronic illness[b]	
Yes	219 (68.2)
No	102 (31.8)
Visits to primary care provider in last 12 months	
1-2	162 (50.6)
3 or more	158 (49.4)

[a]Respondents from UCHealth University of Colorado Hospital and University of Washington Medicine only; we were unable to survey patients at Beth Israel Deaconess Medical Center and Dartmouth-Hitchcock Medical Center in Spring 2020 due to constraints imposed by the COVID-19 pandemic. Table excludes 204 respondents who did not remember completing any previsit forms.

[b]Do you have a chronic illness such as asthma, chronic obstructive pulmonary disease, diabetes, high blood pressure, arthritis, heart disease, or cancer?

More than half of the patients (170/321, 52.9%) completed more than one previsit form (Table 3). In general, few patients reported any difficulty, and more than 80.6% (259/321) had no difficulty answering the questions. About 68.8% (221/321) reported that answering the questions helped them prepare for the visit. Among the 85% (273/321) who reported reading their notes after the visit, 41% (112/273) found that the note included some or all that they had written on the previsit form, whereas more than one-third did not know whether what they had written was included in their visit notes. Patients overwhelmingly agreed that sending their commentary was a good idea (296/321, 92.2%). Most patients preferred to receive the previsit form 2-7 days before their visit. We found no substantive differences in responses between patients who did and did not report having a chronic illness (data not shown).

Table 3. Responses from 321 patients about their experiences with OurNotes.

Question	Responses	Value, n (%)
In the last 12 months, how many times did you complete the OurNotes questionnaire before a visit?		
	Once	151 (57)
	More than once	170 (53)
How difficult was it to answer, "How have you been since your last visit?"		
	Not at all difficult	259 (80.7)
	Somewhat difficult	60 (18.7)
	Very difficult	2 (0.6)
How difficult was it to answer, "What issues would you like to focus on at your visit?"		
	Not at all difficult	286 (89.1)
	Somewhat difficult	32 (10)
	Very difficult	3 (0.9)
Did answering the questions help you prepare for the visit?		
	Yes	221 (68.8)
	No	66 (20.6)
	Don't know	34 (10.6)
Did answering the questions change the conversation between you and your provider?		
	Yes, positive effect	135 (42.1)
	No effect	100 (31.2)
	Yes, negative effect	7 (2.2)
	Don't know	79 (24.6)
Did answering the questions help you and your provider make decisions about your care?		
	Yes, it helped	150 (46.7)
	No effect	90 (28.1)
	No, more difficult	3 (0.9)
	Don't know	78 (24.3)
After the visit, did you read your provider's note?		
	Yes	273 (85)
	No	33 (10.3)
	Don't know	15 (4.7)
Did the provider's note include what you had written? (If yes)		
	Yes, some or all	112 (41)
	No, but clearly provider had read my answers	35 (12.8)
	No, not included	23 (8.4)
	Don't know	103 (37.7)
In general, sending an update about myself before a visit is a good idea.		
	Agree or somewhat agree	296 (92.2)
	Disagree or somewhat disagree	21 (6.5)
	Don't know	4 (1.3)
In general, sending the issues that I want to focus on before visit is a good idea.		
	Agree or somewhat agree	301 (93.8)
	Disagree or somewhat disagree	17 (5.3)
	Don't know	3 (0.9)

Question	Responses	Value, n (%)
When would you like to receive the request for your answers?		
	2 days before the visit	131 (43)
	3-7 days before the visit	161 (53)
	>7 days before the visit	12 (4.0)

The patients wrote 282 comments in free text (Textbox 2). The coders agreed on codes for 161 responses and reached a consensus through discussion on the remaining 121. The comments most often (N=55) addressed the two questions in the form; about half wrote that they liked the questions as they were, and half found the questions too general. More specific comments related to the question asking for an interim history, with many making suggestions for more detailed queries, or requesting guidance on how to respond. The respondents wrote 33 comments about other aspects of the form and suggested some improvements. About half of the respondents commented on the space available for answers thought it adequate, and half wanted more room. Respondents saw the form as a tool supporting their engagement with care (27 comments), most often as a mechanism and incentive for collecting their thoughts before a visit. The 20 comments related to workflow often focused on improved visit efficiency, although several mentioned finding it inefficient when they were asked the same questions again at the visit; 27 comments referenced perceptions that the submitted forms were not read by their clinicians. Several suggested alerts to let the provider know the form was available, and to let the patient know it had been opened. Among other topics mentioned were concerns about privacy, patients for whom the form might present difficulties, and appreciation for specific staff, clinicians, portals, or institutions.

Textbox 2. Examples from 282 patient responses to: “What improvements would you suggest for the future?” and “Please write any other comments you would like to make.”

The 2 questions

- The questions seem alright. It leaves room for the response to be personalized.
- The first question is a bit broad. It wasn't entirely clear to me how much detail my doctor would like or need.
- As someone not familiar with health care, I want more guidance around what is helpful for me to share with my provider about my health. I need examples or clearer descriptions.
- When it says “since last visit”—tell me when the last time I saw this provider was, I see several different doctors...

The form

- I think the form is easy to use, very clear and has just the right amount of space.
- I would not suggest making it any more complicated.
- The shorter the form, the more likely I'll do it.
- It would be helpful to have a brief recap of what our last visit entailed.
- Need more space and the ability to attach documents.

Patient engagement

- I like how it focuses me on what needs to be accomplished.
- Sometimes, I have to sit there and think about more than the first question, but that's ok.
- It leads me to the point that I have superior care in part because of my proactive participation in my own care.
- This program improves doctor-patient communication and facilitates my sense of being connected to the physician.
- I think leaving it so we can fill it in is super helpful. It makes some of the tougher subjects easier to approach.

Workflow

- I found that this makes the flow of the appointment better.
- I prefer to provide the updates online rather than filling out the form when I check in. It seems more efficient for all of us.
- Although, I spend time answering these, it seems I am asked again at the start of my appointment. This is repetitive and unnecessary.
- Some kind of follow up-did the provider open the note? After the visit, did my issues/concerns get addressed?
- Have the doctor have a printout before they enter patient’s room so they can discuss every issue together.

Was the submitted form read?

- I really don't know if it is read before the visit.
- The breakdown was from the doctor, who either had not read them or else for some reason wanted to quiz me to see if I knew what I had written...The time I had spent completing the notes was not treated with respect.
- I did not feel that the questionnaire was helpful, as it was not mentioned or referenced by the doctor or the team.
- I’m not sure the providers saw my answers before the visit. Perhaps they need an alert or flag when the patient has completed these questionnaires?
- It would be nice if I received some confirmation that my notes were received and read.

Other

- I would caution using this method with older people: my 79-year-old mother has difficulty with online medical forms and gets angry or frustrated with them.
- Too many complex medical issues for me to place this in a previsit note.
- I have some privacy issues with typing details of my doctor's concerns into the system in advance. I’m ok typing in the general purpose. I don’t have control over the information once I type it into the system and data tends to stick around forever.
- Let me know if it will be anonymous or who will know.

Discussion

Principal Findings

To our knowledge, this is the first study published on patients and clinicians cogenerating visit notes, or *OurNotes*. Patients were asked to prepare both interim histories about their conditions and their goals for an upcoming scheduled primary care visit, and participating patients and clinicians were strongly supportive. Consistent with our prime hypothesis, patients found completing the forms largely beneficial, and our hypothesis that *OurNotes* would have a neutral or positive impact on the time of clinicians appears to have been substantiated. We did not find additional benefits for patients with chronic illnesses.

In this pilot inquiry, participants offered many suggestions on how to improve the process. Patient respondents described how thinking through and providing the information supported their own engagement with care and helped visits flow more efficiently, but many wrote that they needed guidance about what to write. Clinicians often included the submissions of patients in their visit notes, suggested technical changes that would facilitate the process, and encouraged future efforts to teach patients to write interim histories that are more informative. Overall, both patients and clinicians suggest that this provides an important opportunity for better care. Based on this early proof of concept, all 4 sites decided after the 12-month pilot to continue the intervention, while considering also how to enhance this new approach to transparent patient-clinician communication and patient engagement.

Challenges in the Pilot Implementation

Patients, in aggregate, submitted forms for fewer than 10% of eligible visits. We have no direct evidence from nonparticipants about why they chose not to send forms, and demographic data to analyze differences between them and participants were not available. A number of factors could explain this low level of participation, including insufficient communication with patients about the project before the intervention began, and the stated desire of patients in survey comments for guidance in answering the questions. The factors discussed below may have also contributed to this. Although evaluating reasons for nonparticipation was not among the objectives of this pilot study, it will be important to ascertain this in future research.

Although patients generally liked the idea of submitting information before visits, if such submissions are not readily available to providers before or during the visit, it could prove hurtful, as some patients may feel their input is being overlooked. Another challenge is how to support patients who may feel intimidated by being asked to write free text. Perhaps emphasizing the acceptability of writing phrases rather than complete sentences, allowing dictation functions such as those on smartphones, or offering other ways to record a verbal history may help. For patients to prepare and for clinicians to respond, designing a variety of effective approaches will require input from clinicians, patients, researchers, and information systems specialists.

Incorporating patient-written text into visit notes was an entirely new functionality in all 4 sites, and each designed and implemented an approach within the timeframe of the pilot to enable the most important steps in the process while attempting to minimize the impact on clinicians' workflow. All sites experienced occasional glitches; clinicians commented particularly about sometimes not knowing that a patient had submitted a form until after the visit, and about issues with getting text of patients into their notes. After 12 months, each practice clearly required technical changes to facilitate various steps in the process. Although most clinicians reported no negative impact on their time, this important factor needs further study.

Three Recommendations for Future Implementations

First, both patients and clinicians wanted a *closed communication loop* for previsit information. Clinicians expressed frustration about sometimes not knowing patients had submitted forms, or not being able to find them, and patients were disappointed or frustrated after taking time to complete the forms and then having their clinicians seemingly ignore what they had done. In a closed loop, clinicians would always know before an appointment that the patient had sent information, and patients would know whether the clinician had viewed their submissions. Using automated notifications for both patients and providers, technology can create such a system. Completed forms could be delivered directly to clinicians, although they worry about keeping up with an ever-growing in-box and potential medicolegal liability in unread items. Including links to submitted forms on daily schedules of clinicians could be one way to avoid additional messaging. Providers may also consider deadlines or messaging to encourage early patient submissions, in consideration for clinicians who review patients' data the day or evening before a session, rather than at the time of the visit.

Second, both patients and clinicians explicitly asked for *education* that would help patients write *informative interim histories.* Some patients wrote of simply not knowing how to approach a request for a history. Lacking educational interventions, clinicians found the histories to be less useful than the visit priorities patients formulated. A foundational question concerns whether to ask patients to use a semistructured format, in addition to or in place of free text. In an informal meeting with members of one patient family advisory council, participants suggested that patients should be offered both options. Alternatively, might we prompt for interim histories by problem, or make forms that solicit unstructured histories before primary care visits, whereas those to subspecialists might be better suited to structured queries and responses?

Third, it is clear from the patient participation rate that we must learn more about how to *encourage patients* to take advantage of new opportunities to provide information to their clinicians outside of time-pressured encounters. For many, this involves registering more patients on patient portals, particularly those from vulnerable populations, and ensuring that they have adequate access, education, and support to use the portal [20,21]. Today, many patients fill out forms on paper or tablet computers in the waiting room [22]. Completing a web-based form at home facilitates timely review of prior open notes and may invite more consideration and thought, but it may also raise privacy

concerns. Some survey respondents were intimidated by the form. Many of these concerns can be addressed with education, testimonials by patients and clinicians, and other outreach strategies. An appealing interface, dictation function, and seamless submission process may be helpful. Using some principles of behavioral economics, such as incentives for both patients and clinicians, may also improve rates of participation [23-25]. Encouraging patient engagement is always a multi-pronged, long-term undertaking.

Limitations

This pilot study had important limitations. Most importantly, the proportion of patients who returned previsit forms was modest, and we were unable to conduct postimplementation evaluations with patients in 2 of the 4 sites because of the COVID-19 pandemic. Although we found no difference between the experiences of patients with and without chronic illness, we did not examine records to more fully ascertain their burden of illness. Response rates were within the usual range for web-based surveys [26], but nearly 40% of the patients did not remember submitting a previsit form, a function perhaps of having completed a single form many months before the survey, or of the ever-increasing amount of electronic communication with which patients contend. Finally, we did not collect quantitative information about the proportion of submitted forms that were actually used in the visits.

Conclusions

Overall, in 4 different sites, each with a common goal but many degrees of freedom to design their own intervention, providers and patients liked the idea of patients providing information in advance of their visits and incorporating their observations into visit notes. Further development and research will be needed to assess barriers to patient participation, establish reliable notifications about information submitted by patients, and enable seamless incorporation of patient-generated text into notes.

This exploration is just beginning, and many questions arise. For both patients and clinicians, how should one balance time and effort before an encounter with time spent during a visit? Can cogenerating notes improve a crucial element of care: supporting mutual trust among patients, families, and clinicians? Could cogeneration of notes have particular benefits for those with multiple chronic conditions and those nearing the end of life? Can we effectively bring care partners into the process? Should we ask patients to think more broadly about their goals for health care, addressing issues germane to far more than an individual, time-constrained visit? As care delivery evolves in response to the COVID-19 pandemic, might *OurNotes* prove an effective component of telemedicine [27]? Future exploration may provide useful insights.

Acknowledgments

The authors thank the Commonwealth Fund for primary support and the Gordon and Betty Moore Foundation, Cambia Health Foundation, and Drane Family Fund for further support of this work. The authors gratefully acknowledge colleagues at UCHealth: Kate Sanfillippo, RN, Program Manager, and Matthew Mimnall, Manager of Information Technology, for their work with Epic to enable the technical infrastructure and for building tools and reports that allowed this work to succeed. The authors also thank Jon Darer, MD, MPH, who coined the term '*OurNotes*.' Finally, the authors offer their sincere gratitude to the patients and clinicians who took a chance and participated in the pilot study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Implementation features of OurNotes in four sites.
[DOCX File , 15 KB - jmir_v23i11e29951_app1.docx]

Multimedia Appendix 2
Previsit invitation example.
[DOCX File , 14 KB - jmir_v23i11e29951_app2.docx]

Multimedia Appendix 3
OurNotes previsit form.
[DOCX File , 17 KB - jmir_v23i11e29951_app3.docx]

References

1. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, health, and cost. Health Affairs 2008 May 01;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
2. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Internal Med 2020 Jun 02;172(11 Suppl):123-129. [doi: 10.7326/m19-0876] [Medline: 32479176]

3. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
4. 21st Century Cures Act: Interoperability, information blocking, and the ONC Health IT certification program. Office of the National Coordinator for Health IT, Federal Register. 2020 Mar 19. URL: https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification [accessed 2021-09-09]
5. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
6. New survey data reveals 54 million people are able to access clinicians' visit notes online. OpenNotes®. 2020. URL: https://www.opennotes.org/news/new-survey-data-reveals-54-million-people-are-able-to-access-clinicians-visit-notes-online/ [accessed 2021-09-09]
7. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with Opennotes. Appl Clin Inform 2019 Jan;10(1):10-18. [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
8. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Informatics Assoc 2014 Oct 28:380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
9. Denneson LM, Chen JI, Pisciotta M, Tuepker A, Dobscha SK. Patients' positive and negative responses to reading mental health clinical notes online. Psychiatr Serv 2018 May 01;69(5):593-596. [doi: 10.1176/appi.ps.201700353] [Medline: 29493408]
10. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes After 7 Years: Patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876. [doi: 10.2196/13876]
11. Bell SK, Folcarelli P, Fossa A, Gerard M, Harper M, Leveille S, et al. Tackling ambulatory safety risks through patient engagement: what 10,000 patients and families say about safety-related knowledge, behaviors, and attitudes after reading visit notes. J Patient Saf 2018 Apr 27:1-9. [doi: 10.1097/PTS.0000000000000494] [Medline: 29781979]
12. Bell SK, Folcarelli PH, Anselmo MK, Crotty BH, Flier LA, Walker J. Connecting Patients and Clinicians: The anticipated effects of open notes on patient safety and quality of care. Jt Comm J Qual Patient Saf 2015 Aug;41(8):378-384. [doi: 10.1016/s1553-7250(15)41049-9]
13. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]
14. Bell SK, Gerard M, Fossa A, Delbanco T, Folcarelli PH, Sands KE, et al. A patient feedback reporting tool for OpenNotes: implications for patient-clinician safety and quality partnerships. BMJ Qual Saf 2016 Dec 13:312-322. [doi: 10.1136/bmjqs-2016-006020] [Medline: 27965416]
15. Bourgeois F, Fossa A, Gerard M, Davis ME, Taylor YJ, Connor CD, et al. A patient and family reporting system for perceived ambulatory note mistakes: experience at 3 U.S. healthcare centers. J Am Med Inform Assoc 2019 Dec 01;26(12):1566-1573 [FREE Full text] [doi: 10.1093/jamia/ocz142] [Medline: 31504576]
16. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2016 May 18:262-270. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
17. West C, Dyrbye L, Shanafelt T. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
18. Fulmer T, Reuben DB, Auerbach J, Fick DM, Galambos C, Johnson KS. Actualizing better health and health care for older adults. Health Aff (Millwood) 2021 Feb;40(2):219-225. [doi: 10.1377/hlthaff.2020.01470] [Medline: 33476185]
19. Mafi JN, Gerard M, Chimowitz H, Anselmo M, Delbanco T, Walker J. Patients contributing to their doctors' notes: Insights from expert interviews. Ann Intern Med 2018 Feb 20;168(4):302-305 [FREE Full text] [doi: 10.7326/M17-0583] [Medline: 29132154]
20. Grossman LV, Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
21. Casillas A, Abhat A, Mahajan A, Moreno G, Brown AF, Simmons S, et al. Portals of Change: How patient portals will ultimately work for safety net populations. J Med Internet Res 2020 Oct 23;22(10):e16835 [FREE Full text] [doi: 10.2196/16835] [Medline: 33094732]
22. Anderson MO, Jackson SL, Oster NV, Peacock S, Walker JD, Chen GY, et al. Patients typing their own visit agendas into an electronic medical record: Pilot in a safety-net clinic. Ann Fam Med 2017 Mar;15(2):158-161 [FREE Full text] [doi: 10.1370/afm.2036] [Medline: 28289116]

23. Thaler R, Sunstein C. Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven, Connecticut: Yale University Press; 2008:1-304.
24. Shea J, Adejare A, Volpp K, Troxel AB, Finnerty D, Hoffer K, et al. Patients' views of a behavioral intervention including financial incentives. Am J Manag Care 2017 Jun;23(6):366-371 [FREE Full text] [Medline: 28817301]
25. Jenssen BP, Buttenheim AM, Fiks AG. Using behavioral economics to encourage parent behavior change: Opportunities to improve clinical effectiveness. Acad Pediatr 2019 Jan;19(1):4-10 [FREE Full text] [doi: 10.1016/j.acap.2018.08.010] [Medline: 30172918]
26. NORC at University of Chicago. Methodology report nationwide CAHPS surveys of adults enrolled in Medicaid between October and December 2013. Medicaid.gov. 2015. URL: https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/methodology-report.pdf [accessed 2021-09-09]
27. Kriegel G, Bell S, Delbanco T, Walker J. Covid-19 as Innovation Accelerator: Cogenerating telemedicine visit notes with patients. NEJM Catal 2020 May 20:1-7. [doi: 10.1056/CAT.20.0154]

Abbreviations

BIDMC: Beth Israel Deaconess Medical Center
DHMC: Dartmouth-Hitchcock Medical Center
REDCap: Research Electronic Data Capture
UCH: UCHealth University of Colorado Hospital
UW: University of Washington Medicine

Edited by R Kukafka; submitted 26.04.21; peer-reviewed by C Longhurst, J Hefner, E Blechman; comments to author 18.06.21; revised version received 27.07.21; accepted 12.08.21; published 08.11.21.

Please cite as:
Walker J, Leveille S, Kriegel G, Lin CT, Liu SK, Payne TH, Harcourt K, Dong Z, Fitzgerald P, Germak M, Markson L, Jackson SL, Shucard H, Elmore JG, Delbanco T
Patients Contributing to Visit Notes: Mixed Methods Evaluation of OurNotes
J Med Internet Res 2021;23(11):e29951
URL: https://www.jmir.org/2021/11/e29951
doi:10.2196/29951
PMID:34747710

Original Paper

Determinants of Use of the Care Information Exchange Portal: Cross-sectional Study

Ana Luisa Neves[1,2,3], MSc, MD, PhD; Katelyn R Smalley[1], BA (Hons), MSc; Lisa Freise[1], BSc, MSc; Paul Harrison[4], BSc (Hons), MBA; Ara Darzi[1], PC, KBE; Erik K Mayer[1], BSc, MBBS, MRCS, PhD

[1]Patient Safety Translational Research Centre, Institute of Global Health Innovation, Imperial College, London, United Kingdom

[2]Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal

[3]Department of Community Medicine, Health Information and Decision, University of Porto, Porto, Portugal

[4]Imperial National Health Service Healthcare Trust, London, United Kingdom

Corresponding Author:
Ana Luisa Neves, MSc, MD, PhD
Patient Safety Translational Research Centre
Institute of Global Health Innovation
Imperial College
St Mary's Campus
Queen Elizabeth Queen Mother Wing
London, W2 1NY
United Kingdom
Phone: 44 (0)20 7589 5111
Email: ana.luisa.neves@gmail.com

Abstract

Background: Sharing electronic health records with patients has been shown to improve patient safety and quality of care. Patient portals represent a convenient tool to enhance patient access to their own health care data. However, the success of portals will only be possible through sustained adoption by its end users: the patients. A better understanding of the characteristics of users and nonusers is critical for understanding which groups remain excluded from using such tools.

Objective: This study aims to identify the determinants of the use of the Care Information Exchange, a shared patient portal program in the United Kingdom.

Methods: A cross-sectional study was conducted using a web-based questionnaire. Information collected included age, gender, ethnicity, educational level, health status, postcode, and digital literacy. Registered individuals were defined as having had an account created in the portal, independent of their actual use of the platform; users were defined as having ever used the portal. Multivariate logistic regression was used to model the probability of being a user. Statistical analysis was performed in R and Tableau was used to create maps of the proportion of Care Information Exchange users by postcode area.

Results: A total of 1083 participants replied to the survey (186% of the estimated minimum target sample). The proportion of users was 61.58% (667/1083). Among these, most (385/667, 57.7%) used the portal at least once a month. To characterize the system's users and nonusers, we performed a subanalysis of the sample, including only participants who had provided at least information regarding gender and age. The subanalysis included 650 individuals (389/650, 59.8% women; 551/650, 84.8% >40 years). Most participants were White (498/650, 76.6%) and resided in London (420/650, 64.6%). Individuals with a higher educational degree (undergraduate and professional, or postgraduate and higher) had higher odds of being a portal user (adjusted odds ratio [OR] 1.58, 95% CI 1.04-2.39 and OR 2.38, 95% CI 1.42-4.02, respectively) compared with those with a secondary degree or below. Higher digital literacy scores (≥30) were associated with higher odds of being a user (adjusted OR 2.96, 95% CI 2.02-4.35). Those with a good overall health status had lower odds of being a user (adjusted OR 0.58, 95% CI 0.37-0.91).

Conclusions: This work adds to the growing body of evidence highlighting the importance of educational aspects (educational level and digital literacy) in the adoption of patient portals. Further research should not only describe but also systematically address these inequalities through patient-centered interventions aimed at reducing the digital divide. Health care providers and policy makers must partner in investing and delivering strategic programs that improve access to technology and digital literacy in an effort to improve digital inclusion and reduce inequities in the delivery of care.

(J Med Internet Res 2021;23(11):e23481) doi:10.2196/23481

KEYWORDS

patient portals; electronic health records; patient participation

Introduction

Background

A growing body of evidence supports providing patients access to their electronic health records to improve several aspects of quality of care, including patient safety [1-3], patient-centeredness [4,5], and effectiveness [3]. Patient portals are currently recognized as a promising mechanism for improving health care data sharing with patients. Patients may use portals for a range of purposes, including entering, retrieving, or sharing their health care information, communicating with health care providers, and self-managing their health [6]. The use of patient portals can improve health outcomes (eg, in the case of type 2 diabetes) [7,8], increase patient satisfaction [9], improve medication safety and adherence [10-12], and improve communication between the patient and the health care provider [9-13].

The success of patient portals and the subsequent achievement of their proposed benefits will only be possible through sustained adoption by its end users: the patients. However, questions remain about how health care providers and policy makers can encourage sustainable adoption by patients without exacerbating the pre-existing digital divide or widening discrepancies in the delivery of care [14]. In fact, despite the increasing implementation of patient portals by health care institutions and governments worldwide, adoption by patients has remained slower than expected [15,16]. A meta-analysis published by Fraccaro et al [17], including 40 studies, showed an overall mean adoption rate of 52% (95% CI 42%-62%). However, the authors emphasize that the evaluation of adoption in clinical practice may have different results from those obtained in randomized controlled trials [17].

Several individual and sociotechnical factors have been suggested to affect portal adoption, such as age, health status, educational level, and patient activation (ie, the knowledge, skills, and confidence a person has in managing their own health and care) [18-20]. Although there is some evidence of higher adoption by those with poorer health status [21,22] and higher educational level [20], there is mixed evidence about the impact of age [21,23] and patient activation [24,25]. Technology-related factors may also play an important role, with higher digital health literacy, better portal design, and higher perceived usefulness and ease of use being potentially associated with a positive impact [26].

In 2015, Patients Know Best teamed up with Imperial College National Health Service Healthcare Trust to roll out the *Care Information Exchange* (CIE) across North West London the largest shared patient portal program in the United Kingdom, hosting records of more than 2.3 million people living in North West London [27]. The CIE collects data from hospitals and general practitioners' practices in North West London and 15 other hospitals outside of North West London, including Birmingham, Bristol, Liverpool, Manchester, Scotland, and Wales [27]. The CIE was the first to introduce *mass registration*, enabling people to sign up and access their health record at scale and with speed in a number of ways: either by speaking to a member of staff, using the kiosk check-in screen commonly found in hospital outpatient waiting rooms, or by letter of invitation to their home. The CIE contains patient information, including appointment details, test results, care plans, discharge summaries, clinical letters, and information on medications. If a patient's primary care practice has signed up, data such as allergies, medications, and diagnoses will also be visible to them. Patients may access their records whenever they wish to review information or when notified about new information, such as available test results [3].

Objectives

This study aims to characterize individuals registered with the CIE and explore the differences between users and nonusers in terms of their demographic, geographic, health status, and educational characteristics (ie, educational level and digital health literacy) and motivation to be involved in their own health care (as a proxy measure for patient activation), thus identifying the main determinants of use of the portal. Our hypothesis is that the users' characteristics described above can affect the adoption of CIE. This is key to understanding barriers to adoption as well as understanding which groups remain underserved or excluded from using patient portals, which is critical for future patient-centered digital health care delivery.

Methods

Study Design, Participants, and Data Collection

We conducted a cross-sectional study using an anonymous web-based questionnaire presented by Qualtrics.

Patients registered with the CIE portal and who had logged in at least once during the study period (n=27,411) were invited to follow the link to complete the survey. There were no specific exclusion criteria; however, patients needed to be ≥18 years to register to use the portal. This link contained general information about the purpose of the survey, and informed consent was obtained at the beginning of the survey. Considering this population, a confidence level of 95% and a margin of error of 5%, the minimum sample size to ensure representativeness was calculated as n=379 respondents. The survey was open for completion between July 1, 2018, and July 1, 2019. No patient identifiers were collected. Information collected included age, gender, ethnicity, educational level, postcode (first part), digital literacy, health status, and motivation to be involved in their own health care (as a proxy for patient activation).

Measurements

Age was categorized into age bands (<30, 31-40, 41-50, 51-65, ≥65), and ethnicity was categorized as White or Black, Asian, and minority. The first part of the postcode was categorized as

London's official postal district for descriptive purposes [28]. For the univariate and multivariate analyses, owing to the highly skewed distribution toward West and North West London, postcode areas were categorized as West London, North West London, other London, or other.

Digital literacy was assessed using the eHealth Literacy Scale (eHEALS), developed by Norman and Skinner [29]. This tool identifies six core skills or literacies: (1) traditional literacy, (2) health literacy, (3) information literacy, (4) scientific literacy, (5) media literacy, and (6) computer literacy. On the basis of these core literacies, the eHEALS tool assesses consumers' knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information to health problems. The eHEALS tool uses a 5-point Likert scale (1-strongly disagree and 5-strongly agree), with a score ranging from 8 to 40, with a higher score indicating higher literacy.

Overall health status was assessed via a multiple-choice question ("How good do you think your health is?" with possible responses: "Very good," "Somewhat good," "Neither good nor poor," "Somewhat poor," and "Very poor"). Motivation to be involved in their own care was similarly assessed via multiple-choice questions ("In general, how motivated to be involved in your health care are you?" with possible responses: "A little," "A moderate amount," "A lot," and "Very much").

Registered individuals were defined as having had an account created in the CIE portal, independent of their actual use of the platform. *Users* and *nonusers* were defined as individuals having answered "Yes" or "No," respectively, to the question "Have you ever used CIE?" Those who answered "Yes" (ie, users) were also asked about their frequency of use ("How often do you use CIE?" with response options as follows: "Less than once a month," "Once a month," "Once a week," and "Twice a week or more").

Data Analysis

Mean and SD were calculated for continuous variables, and proportions and counts were calculated for categorical variables. Univariate logistic regression modeled the odds of being a user as a function of each individual predictor. The resulting coefficients, expressed as log (odds) ratios, were transformed into crude odds ratios (ORs) with a 95% CI.

Multivariate logistic regression was used to model the probability of being a user as a function of age, gender, educational level, digital literacy (categorical variables), and overall health status. The variables were chosen for multivariate analysis through automated, backward stepwise elimination. With this procedure, all variables of interest are included in the first iteration of the model and removed one by one, starting with the ones for which elimination would improve the model fit most and ending the process when removing an additional variable worsens the model fit. Model quality comparisons were conducted using the Akaike information criterion [30]. Basic demographic variables (age and gender) were inputted as forced-in covariates in the multivariate analysis. Adjusted ORs with 95% CI were calculated.

Statistical analyses were conducted in RStudio, using the *plyr*, *dplyr*, *ggplot2*, and *car* packages. Tableau software was used to create maps of the total number of participants and the proportion of CIE users by postcode area.

Ethics

The study was approved as a Service Evaluation at Imperial College Healthcare National Health Service Trust (Registration Number: 296/2018).

Results

Participants' Characteristics

The survey link was shared with a total of 27,411 patients that logged at least once (ie, were accredited to use the system) between July 1, 2018, and July 1, 2019. A total of 1083 subjects replied to the survey (186% of the estimated target sample). The proportion of users was 61.58% (667/1083), and among these, more than half (385/667, 57.7%) used the portal at least once a month. Self-identified users and nonusers of CIE were more likely to provide demographic information (for age: 251/667, 37.6% and 152/416, 36.5%, respectively; for gender: 246/667, 36.9%) and 36.6% (152/416). Of them, 650 participants provided information regarding their gender and age, and we limited the analysis to these individuals (+71.5% of the estimated target sample).

In the subanalysis of patients who provided basic characteristics regarding gender and age category, 59.8% (389/650) were women, and 84.8% (551/650) were ≥40 years. Most participants were White (498/650, 76.6%) and resided in London (651/1006, 64.7%). Among them, 55.9% (363/651) were from North West London. A more detailed overview of the distribution of participants by postcode area is provided in Figure 1. The mean literacy score assessed by the eHEALS tool was 31.5 (SD 7.9), and 22.3% (145/651) had a postgraduate degree or higher. Most participants considered themselves very motivated to be involved in their own care (374/651, 57.5%), and 41.8% (272/651) considered themselves to have a good or very good health status. A full description of the analyzed sample and the characteristics of the nonuser and user groups is provided in Table 1.

Figure 1. Geographic location overview. General overview of England (left) and Central London (right, representing 64.7% of the subjects). Circle size represents the total number of respondents per postcode area, and color code represents the percentage of Care Information Exchange users per postcode area. The right-side image shows the stronger representation of North West London in the sample. CIE: Care Information Exchange.

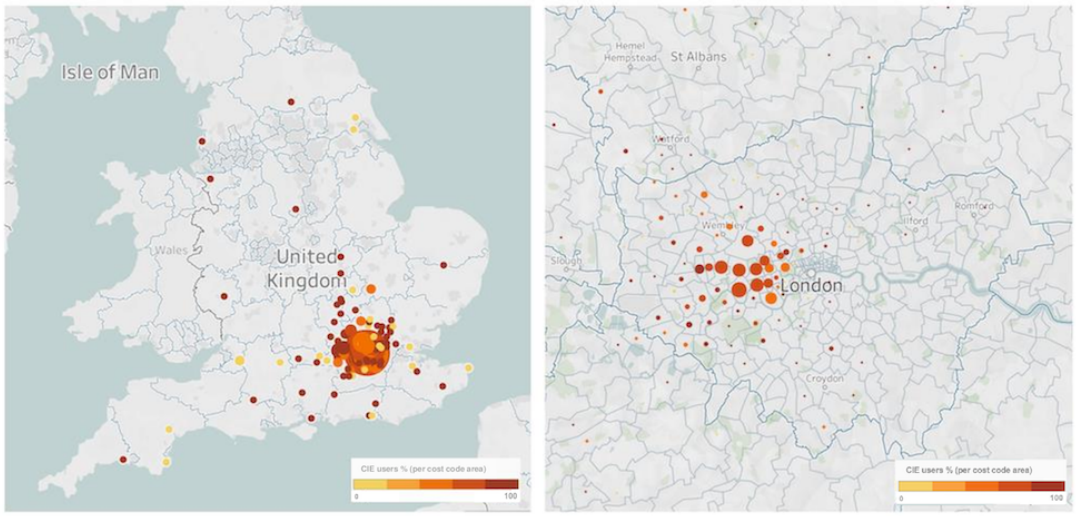


Table 1. Characteristics of the participants according to their use of the system (N=650).

Characteristics	Nonusers (n=205)	Users (n=447)	Total (N=650)
Gender, n (%)			
Female	113 (55.1)	276 (61.7)	389 (59.8)
Male	91 (44.4)	167 (37.4)	258 (39.7)
Other	1 (0.5)	2 (0.4)	3 (0.5)
No response	0 (0)	0 (0)	0 (0)
Age band (years), n (%)			
<30	9 (4.4)	22 (4.9)	31 (4.8)
31-40	20 (9.8)	48 (10.7)	68 (10.5)
41-50	23 (11.2)	62 (13.9)	85 (13.1)
51-65	72 (35.1)	166 (37.1)	238 (36.6)
≥65	81 (39.5)	147 (32.9)	228 (35.1)
No response	0 (0)	0 (0)	0 (0)
Ethnicity, n (%)			
BAME[a]	34 (16.6)	75 (16.8)	109 (16.8)
White	155 (75.6)	343 (76.7)	498 (76.6)
Other	16 (7.8)	22 (4.9)	38 (5.8)
No response	0 (0)	5 (1.1)	5 (0.8)
Geographic location, n (%)			
London			
East	2 (0.9)	2 (0.5)	4 (0.6)
East Central	0 (0)	0 (0)	0 (0)
North	0 (0)	7 (1.6)	7 (1.1)
North West	15 (7.3)	40 (8.9)	55 (8.5)
South East	0 (0)	5 (1.1)	5 (0.7)
South West	20 (9.8)	28 (6.3)	48 (7.4)
West	96 (46.8)	203 (45.4)	299 (46)
West Central	1 (0.5)	1 (0.2)	2 (0.3)
Other	62 (30.2)	146 (32.7)	208 (31.8)
No response	9 (4.4)	13 (2.9)	22 (3.5)
Educational degree, n (%)			
Secondary school or below	75 (36.6)	118 (61.1)	193 (29.7)
Undergraduate or professional degree	77 (37.6)	180 (40.3)	257 (39.5)
Postgraduate or higher	33 (16.1)	112 (25.1)	145 (22.3)
No response	20 (9.8)	35 (7.8)	55 (8.5)
Digital literacy (eHEALS[b] score), mean (SD)	28.4 (8.1)	32.9 (7.4)	31.5 (7.9)
Overall health status, n (%)			
Good or very good	95 (46.3)	177 (39.6)	272 (41.8)
Neither good nor poor	55 (26.8)	106 (23.7)	161 (24.8)
Poor or very poor	55 (26.8)	162 (36.2)	217 (33.3)
No response	0 (0)	0 (0)	0 (0)
Motivation to be involved in own care, n (%)			

Characteristics	Nonusers (n=205)	Users (n=447)	Total (N=650)
Not very much	7 (3.4)	6 (1.34)	13 (2)
A moderate amount	40 (19.5)	43 (9.6)	83 (12.7)
A lot	61 (29.8)	116 (25.9)	177 (27.2)
Very much	96 (46.8)	278 (62.2)	374 (57.5)
No response	1 (0.5)	2 (0.5)	3 (0.5)

[a]BAME: Black, Asian, and minority ethnic.

[b]eHEALS: eHealth Literacy Scale.

Differences Between Users and Nonusers

The characteristics of both users and nonusers were explored using univariate logistic regression (crude ORs) and a logistic regression model with predictors (adjusted ORs). The differences between groups are shown in Table 2.

Crude ORs showed that individuals with a higher educational degree (undergraduate or professional or postgraduate or higher) had higher odds of being a portal user (crude OR 1.48, 95% CI 1.00-2.20 and crude OR 2.15, 95% CI 1.33-3.05, respectively) than those with a secondary degree or below. Higher digital literacy scores (>30) were also associated with higher odds of being a user (crude OR 2.90, 95% CI 2.06-4.11) and those that reported being “very much motivated to be involved in their own care” (crude OR 3.38, 95% CI 1.10-10.3). Participants with a good overall health status had lower odds of being users (crude OR 0.63, 95% CI 0.43-0.94).

Adjusted ORs represent the multivariate analysis of the predictors of CIE use. Initially, all variables were included in the multivariate model, and backward stepwise elimination was used to select the best-fit model. Digital literacy, education, and health status remained in the naive best-fit regression, and gender and age were reinputted as forced-in covariates, as previously described in the *Methods* section. Sensitivity analyses showed minimal differences as a result of their inclusion or exclusion. All covariates that were statistically significant in the naive model remained, and no additional variables gained significance.

Adjusted ORs showed that individuals with a higher educational degree (undergraduate or professional or postgraduate or higher) had higher odds of being a portal user (adjusted OR 1.58, 95% CI 1.04-2.39 and adjusted OR 2.38, 95% CI 1.42-4.02, respectively) than those with a secondary degree or below. Higher digital literacy scores (≥30) were also associated with higher odds of being a user (adjusted OR 2.96, 95% CI 2.02-4.35). Those with a good overall health status had lower odds of being a user (adjusted OR 0.58, 95% CI 0.37-0.91).

A significant association was found with *increased motivation to be involved in own care* (crude OR) for those very motivated to be involved in their own care is 3.38 (95% CI 1.10-10.3). However, it was not possible to explore the effect in multivariate analysis as the variable was removed from the best-fit model as part of the stepwise backward elimination procedure. No significant associations were found with age, gender, ethnicity, or geographic location.

Table 2. Characteristics of users according to their input with crude and adjusted odds ratios (ORs; N=650).

Characteristics	Nonadjusted model[a]		Adjusted model[b]	
	Crude OR (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
Gender				
Female	Reference	N/A[c]	Reference	N/A
Male	0.75 (0.54-1.05)	.01	0.92 (0.624-1.35)	.67
Other	0.81 (0.07-9.12)	.87	0 (0-infinity)	.98
Age band				
<30	Reference	N/A	Reference	N/A
31-40	0.98 (0.39-2.50)	.97	0.63 (0.22-1.76)	.37
41-50	1.10 (0.44-2.74)	.83	0.88 (0.32-2.40)	.80
51-65	0.94 (0.41-2.15)	.89	0.85 (0.34-2.12)	.73
≥65	0.74 (0.33-1.69)	.47	0.65 (0.26-1.65)	.37
Ethnicity[d]				
White	Reference	N/A	__[e]	—
BAME[f] or other	0.88 (0.59-1.33)	.55	—	—
Geographic location[d]				
West London	0.94 (0.65-1.36)	.74	—	—
North West London	1.19 (0.62-2.29)	.41	—	—
Other London	0.85 (0.47-1.53)	.59	—	—
Other	Reference	N/A	—	—
Educational degree				
Secondary or below	Reference	N/A	Reference	N/A
Undergraduate or professional	1.48 (1.00-2.20)	.049	1.58 (1.04-2.39)	.001
Postgraduate or higher	2.15 (1.33-3.50)	.002	2.38 (1.42-4.02)	.03
Digital literacy				
Literacy score <30	Reference	N/A	Reference	N/A
Literacy score ≥30	2.90 (2.06-4.11)	<.001	2.96 (2.02-4.35)	<.001
Overall health status				
Poor	Reference	N/A	Reference	N/A
Neutral	0.65 (0.42-1.02)	.06	0.73 (0.45-1.20)	.21
Good	0.63 (0.43-0.94)	.02	0.58 (0.37-0.91)	.02
Motivation to be involved in own care[d]				
Not very much	Reference	N/A	—	—
A moderate amount	1.25 (0.39-4.05)	.17	—	—
A lot	2.22 (0.71-6.89)	.71	—	—
Very much	3.38 (1.10-10.3)	.03	—	—

[a]Crude odds ratios calculated from univariate logistic regression, where the probability of being a user was modeled.

[b]Logistic regression model with predictors: age, gender, education level, digital literacy, and health status.

[c]N/A: not applicable.

[d]These variables were removed from multivariate analysis using a stepwise backward elimination procedure.

[e]Not available.

[f]BAME: Black, Asian, and minority ethnic.

Discussion

Principal Findings

Participants with an undergraduate or professional degree were 58% more likely to use the portal than those with secondary education or below (adjusted OR 1.58, 95% CI 1.04-2.39), and those with a postgraduate degree were more than 2 times as likely to use the portal (adjusted OR 2.38, 95% CI 1.42-4.02). People with an eHEALS digital literacy score greater than 30 were nearly three times more likely to be portal users than those with eHEALS scores below 30 (adjusted OR 2.96, 95% CI 2.02-4.35). These results reveal the impact of education and literacy on the adoption of digital technologies and reinforce their role as drivers of patient exclusion.

Participants with good overall health status were about half as likely to have reported using the portal compared with those reporting poor health status (adjusted OR 0.58, 95% CI 0.37-0.91). This reinforces that although the ability to use digital technologies is an important contributor to their adoption, the perceived need for service is equally important. People with poor health will have more incentive to engage with technologies that facilitate their health care than those without health challenges.

No significant associations were found with age, gender, ethnicity, geographic location, or motivation to be involved in their own care. The fact that demographic factors such as these were not strongly associated with use further indicates that ability and need to use digital health management tools are potentially the key drivers of their uptake.

Comparison With Previous Literature

Our findings are consistent with previous evidence suggesting that portal users, compared with nonusers, are more often highly educated and have higher eHealth literacy levels [31].

The spread and scale of digitally enabled care are happening fast—in fact, faster than our ability to ensure that all patient groups have the basic digital literacy tools to fully exploit its potential. The educational level seems to be an independent predictor of portal use; in an inpatient study, after adjusting for age, gender, race and ethnicity, immigration status, educational attainment, and employment status, those without an education degree had higher odds of never logging on to the portal [20].

Previous studies have also found that patients with higher eHealth literacy levels are more likely to be portal users [31,32]. Importantly, people's self-perceived skills to use web-based information actually have an impact on their health and the quality of care received, and a lack of such skills may result in adverse health outcomes [33,34]. According to Holt et al [35], information about patients' health literacy may provide a better understanding of patients' reasons for not using digital health services rather than sociodemographic data.

The educational level seems to be an independent predictor of portal use; in an inpatient study, after adjustment for age, gender, race and ethnicity, immigration status, educational attainment, and employment status, those without an education degree had higher odds of never logging on to the portal [20]. Previous studies also found that patients with higher eHealth literacy have a higher likelihood of being portal users [31,32]. Similar findings were reported by Holt et al [35], suggesting that information about patients' health literacy may provide a better understanding of patients' reasons for not using digital health services rather than sociodemographic data.

The association between having a good overall health status and a lower likelihood of being a portal user has also been documented in previous studies. People with disabilities, chronic conditions, and frequent use of health care services (and caregivers of elderly parents or children) tend to be associated with higher patient portal interest and use [19,22]. In this study, we did not find any significant associations with age, gender, ethnicity, or geographic location. The association between age and portal use has been inconsistently reported, and although some studies have suggested that elderly people use portals less often [19,22,23], others did not find a significant effect [36]. Mixed results have also been found regarding gender differences [33,34]. It has been previously suggested that ethnic minorities use patient portals less often [37]. However, a study evaluating disparities in enrollment and use of a patient portal concluded that although minority patients were less likely to register to use a patient portal, there were no racial and ethnic disparities in the use of the patient portal among enrollees, suggesting that the digital divide may be particularly important at enrollment, rather than in continued use (ie, postenrollment) [21]. It is likely that the association between ethnicity and portal use results from a complex relationship modeled by a range of sociodemographic, economic, and educational variables. In this study, ethnicity was not, per se, an independent predictor of portal use.

A significant association was found between portal use and patient activation (expressed as the subjective motivation to be involved in one's own care), but significance did not remain in multivariate analyses—in fact, this variable was removed from the best-fit model as part of the stepwise backward procedure. A few studies exploring this aspect have found inconsistent results: while one study found slightly higher patient activation measure scores in portal users [24], others found no significant associations between patient activation measure levels and portal log-in [25].

Strengths and Limitations

This study had several strengths. The sample size was 75% higher than the estimated minimum sample size to ensure representativeness and adequate statistical power. A comprehensive set of characteristics was collected and analyzed at the individual level, allowing us to explore not only the classic demographic factors (age, gender, ethnicity, and educational level) but also important variables such as overall health status, motivation to self-manage, and health literacy (using a validated tool). The high response rate and the overall large sample size contribute to the robustness of these findings.

Some limitations of this study should also be acknowledged. First, it must be noted that, although we achieved the minimum sample size required, the overall response rate was low, which has important considerations for generalizations about which determinants drive portal adoption. Intrinsically to the study

design, a range of selection biases cannot be excluded. Although web-based surveys are a well-accepted method for data collection, they induce a selection bias by excluding less tech-savvy individuals, individuals with less digital literacy, with less consistent access to the internet, and therefore those who are less likely to adopt patient portals. Using an exclusively web-based recruitment strategy also introduces an additional selection bias, but unfortunately, we were not able to email participants directly because of information governance limitations. In addition, both users and nonusers were registered at the portal, and therefore our results highlight potential determinants of use among registered users and not general determinants of initial engagement with the portal. This needs to be considered in any attempt to perform external generalizations. Although this study aimed to identify the determinants of use between those that had already registered (not the determinants to engage or register with a portal in the first instance), future research should also address determinants of initial engagement (ie, register with a portal in the first instance).

In this study, patient portal use was patient-reported; therefore, a potential information bias could also be present. As an alternative, patient log-in can be used to measure portal use [34]; however, this approach lacks contextual information.

It is also important to note that participants included in this study were predominantly from a specific geographic location (North West London); therefore, these results need to be carefully interpreted in any attempt to perform external generalizations (ie, to other geographic locations, populations, or health care systems). In future work, it would be important to evaluate not only the geographic location of the users but also whether users live in an urban, rural, or mixed setting, given the variation in accessibility (ie, internet access and connectivity options) among those.

Finally, this study aimed to evaluate the impact of individual factors. However, there are a plethora of sociotechnical factors (including factors such as social determinants, portal design, and communication strategies) that may equally influence adoption rates and the impact that is important for evaluation in future research.

Conclusions

This work adds to the growing body of evidence highlighting the importance of educational aspects (educational level and digital literacy) for sustainable implementation and use of patient-facing electronic health record portals. To ensure that all patients are able to benefit from patient portals, it is critical that we move from identifying disparities in portal use to systematically addressing them through patient-centered interventions that reduce the digital divide.

Further research evaluating the impact of interventions to improve portal use must therefore explore the effect on potential disparities in use, addressing the impact on patients with a low educational level, poor access to technology, or lack of ability or confidence to use it for health-related purposes.

Equally, portal use can be improved by co-designing portals with patients, incorporating user-centered design techniques, and ensuring that a diverse group of potential users is included in the process. In particular, involving older persons and those with lower general health literacy and digital health literacy in digital development can provide important insights into the barriers experienced by these typically excluded groups and co-design strategies to overcome them [38].

Therefore, it is critical to ensure that health care providers and policy makers align across sectors, investing and delivering strategic programs that improve access to technology and digital literacy, in an effort to improve digital inclusion and reduce inequities in the delivery of care.

Acknowledgments

This research was funded by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre and supported by the NIHR Imperial Biomedical Research Centre. The research was enabled by the Imperial Clinical Analytics Research and Evaluation environment and used the Imperial Clinical Analytics Research and Evaluation team and data resources. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, Martin T, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014 Feb;21(e1):157-162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
2. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec;29(12):1019-1032 [FREE Full text] [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
3. Freise L, Neves AL, Flott K, Harrison P, Kelly J, Darzi A, et al. Assessment of patients' ability to review electronic health record information to identify potential errors: cross-sectional web-based survey. JMIR Form Res 2021 Feb 26;5(2):e19074 [FREE Full text] [doi: 10.2196/19074] [Medline: 33635277]

4. Dalal A, Dykes P, Collins S, Lehmann LS, Ohashi K, Rozenblum RM, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87 [FREE Full text] [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
5. Maher M, Kaziunas E, Ackerman M, Derry H, Forringer R, Miller K, et al. User-centered design groups to engage patients and caregivers with a personalized health information technology tool. Biol Blood Marrow Transplant 2016 Feb;22(2):349-358 [FREE Full text] [doi: 10.1016/j.bbmt.2015.08.032] [Medline: 26343948]
6. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020 Jun 02;172(11 Suppl):123-129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
7. Sun R, Burke LE, Korytkowski MT, Saul MI, Li D, Sereika SM. A longitudinal examination of patient portal use on glycemic control among patients with uncontrolled type 2 diabetes. Diabetes Res Clin Pract 2020 Dec;170:108483. [doi: 10.1016/j.diabres.2020.108483] [Medline: 33038473]
8. Tenforde M, Jain A, Hickner J. The value of personal health records for chronic disease management: what do we know? Fam Med 2011 May;43(5):351-354 [FREE Full text] [Medline: 21557106]
9. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
10. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
11. Chrischilles E, Hourcade J, Doucette W, Eichmann D, Gryzlak B, Lorentzen R, et al. Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 2014;21(4):679-686 [FREE Full text] [doi: 10.1136/amiajnl-2013-002284] [Medline: 24326536]
12. Sarkar U, Lyles C, Parker M, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
13. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]
14. Graetz I, Gordon N, Fung V, Hamity C, Reed M. The digital divide and patient portals: internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
15. 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]
16. Greenhalgh T, Morris L, Wyatt JC, Thomas G, Gunning K. Introducing a nationally shared electronic patient record: case study comparison of Scotland, England, Wales and Northern Ireland. Int J Med Inform 2013 May;82(5):125-138. [doi: 10.1016/j.ijmedinf.2013.01.002] [Medline: 23434362]
17. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. Patient portal adoption rates: a systematic literature review and meta-analysis. Stud Health Technol Inform 2017;245:79-83. [Medline: 29295056]
18. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006;2006:514-518 [FREE Full text] [Medline: 17238394]
19. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
20. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
21. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1(Suppl 1):8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
22. Ronda MC, Dijkhorst-Oei L, Gorter KJ, Beulens JW, Rutten GE. Differences between diabetes patients who are interested or not in the use of a patient web portal. Diabetes Technol Ther 2013 Jul;15(7):556-563 [FREE Full text] [doi: 10.1089/dia.2013.0023] [Medline: 23777369]
23. Turner K, Clary A, Hong Y, Tabriz AA, Shea CM. Patient portal barriers and group differences: cross-sectional national survey study. J Med Internet Res 2020 Sep 17;22(9):e18870 [FREE Full text] [doi: 10.2196/18870] [Medline: 32940620]
24. Nagykaldi Z, Aspy CB, Chou A, Mold JW. Impact of a wellness portal on the delivery of patient-centered preventive care. J Am Board Fam Med 2012;25(2):158-167 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110130] [Medline: 22403196]
25. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]

26. Lazard A, Watkins I, Mackert M, Xie B, Stephens K, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):157-161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
27. Care information exchange – the largest shared patient portal program in the UK. Patients Know Best. 2019. URL: https://patientsknowbest.com/2019/10/21/care-information-exchange-the-largest-shared-patient-portal-program-in-the-uk/ [accessed 2021-09-01]
28. ONS postcode directory. Office for National Statistics. 2020. URL: http://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-may-2020 [accessed 2020-08-13]
29. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
30. Boisbunon A, Canu S, Fourdrinier D, Strawderman W, Wells MT. Akaike's information criterion, Cp and estimators of loss for elliptically symmetric distributions. Int Stat Rev 2014 Dec;82(3):422-439 [FREE Full text]
31. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015 Nov 5;2015:1871-1880 [FREE Full text] [Medline: 26958286]
32. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
33. Sanders M, Winters P, Fortuna R, Mendoza M, Berliant M, Clark L, et al. Internet access and patient portal readiness among patients in a group of inner-city safety-net practices. J Ambul Care Manage 2013;36(3):251-259. [doi: 10.1097/JAC.0b013e31829702f9] [Medline: 23748274]
34. Jung C, Padman R. Virtualized healthcare delivery: understanding users and their usage patterns of online medical consultations. Int J Med Inform 2014 Dec;83(12):901-914. [doi: 10.1016/j.ijmedinf.2014.08.004] [Medline: 25193501]
35. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: an exploratory survey of a group of medical outpatients. Interact J Med Res 2019 Apr 05;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
36. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
37. Smith S, O'Conor R, Aitken W, Curtis L, Wolf M, Goel M. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
38. Neves AL, Lawrence-Jones A, Naar L, Greenfield G, Sanderson F, Hyde T, et al. Multidisciplinary teams must work together to co-develop inclusive digital primary care for older people. Br J Gen Pract 2020 Nov 26;70(701):582. [doi: 10.3399/bjgp20X713645] [Medline: 33243904]

Abbreviations

CIE: Care Information Exchange
eHEALS: eHealth Literacy Scale
NIHR: National Institute for Health Research
OR: odds ratio

Edited by R Kukafka, G Eysenbach; submitted 14.08.20; peer-reviewed by AV Versluis, Z Reis, A Kapoor; comments to author 17.11.20; revised version received 11.03.21; accepted 09.08.21; published 11.11.21.

Please cite as:
Neves AL, Smalley KR, Freise L, Harrison P, Darzi A, Mayer EK
Determinants of Use of the Care Information Exchange Portal: Cross-sectional Study
J Med Internet Res 2021;23(11):e23481
URL: https://www.jmir.org/2021/11/e23481
doi:10.2196/23481
PMID:34762063

Review

Security Engineering of Patient-Centered Health Care Information Systems in Peer-to-Peer Environments: Systematic Review

Imrana Abdullahi Yari[1], MSc; Tobias Dehling[2,3], PhD; Felix Kluge[1], PhD; Juergen Geck[4], MSc; Ali Sunyaev[2,3], PhD; Bjoern Eskofier[1], PhD

[1]Department of Artificial Intelligence in Biomedical Engineering, Machine Learning and Data Analytics Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

[2]Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany

[3]KASTEL Security Research Labs, Karlsruhe, Germany

[4]REFINIO GmbH, Rohr, Germany

Corresponding Author:
Imrana Abdullahi Yari, MSc
Department of Artificial Intelligence in Biomedical Engineering, Machine Learning and Data Analytics Lab
Friedrich-Alexander University Erlangen-Nuremberg
Carl-Thiersch-Straße 2b
Erlangen, 91052
Germany
Phone: 49 9131 85 20288
Email: imrana.yari.abdullahi@fau.de

Abstract

Background: Patient-centered health care information systems (PHSs) enable patients to take control and become knowledgeable about their own health, preferably in a secure environment. Current and emerging PHSs use either a centralized database, peer-to-peer (P2P) technology, or distributed ledger technology for PHS deployment. The evolving COVID-19 decentralized Bluetooth-based tracing systems are examples of disease-centric P2P PHSs. Although using P2P technology for the provision of PHSs can be flexible, scalable, resilient to a single point of failure, and inexpensive for patients, the use of health information on P2P networks poses major security issues as users must manage information security largely by themselves.

Objective: This study aims to identify the inherent security issues for PHS deployment in P2P networks and how they can be overcome. In addition, this study reviews different P2P architectures and proposes a suitable architecture for P2P PHS deployment.

Methods: A systematic literature review was conducted following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines. Thematic analysis was used for data analysis. We searched the following databases: IEEE Digital Library, PubMed, Science Direct, ACM Digital Library, Scopus, and Semantic Scholar. The search was conducted on articles published between 2008 and 2020. The Common Vulnerability Scoring System was used as a guide for rating security issues.

Results: Our findings are consolidated into 8 key security issues associated with PHS implementation and deployment on P2P networks and 7 factors promoting them. Moreover, we propose a suitable architecture for P2P PHSs and guidelines for the provision of PHSs while maintaining information security.

Conclusions: Despite the clear advantages of P2P PHSs, the absence of centralized controls and inconsistent views of the network on some P2P systems have profound adverse impacts in terms of security. The security issues identified in this study need to be addressed to increase patients' intention to use PHSs on P2P networks by making them safe to use.

(J Med Internet Res 2021;23(11):e24460) doi:10.2196/24460

KEYWORDS

patient-centered; health care; information infrastructures; decentralization; mobile health; peer-to-peer; COVID-19 proximity trackers; edge computing; security; vulnerabilities; attacks; threats; mobile phone

Introduction

Motivation

Patients require access to their health information with the same ease as with other web-based activities such as banking or shopping; however, patients are often only one part of the current health care processes and not the focus of attention [1]. Such limitations of traditional health care processes, widespread individual adoption of digital systems, and advancements in health care practice create a growing demand for patient-centered health care information systems (PHSs). PHSs are scalable information systems that leverage information technology to support patients in managing and taking an active role in their own health [1,2]. PHSs are not designed to replace traditional health care information systems, such as electronic health records, but rather to complement them [3] by offering additional functionalities, such as translation of clinical information into layman's terms [4], provision of information on medications a patient is taking [2,5], or provision of vetted information to support self-administered interventions (eg, reduce weight or quit smoking) [4].

The diversity and flexibility of PHSs enable them to provide any functionality that patients find helpful [2], including maintaining personal health records (PHRs) [6], tracking mental wellness [7], subscribing to risk prediction services for chronic diseases [6,8], and calculating pregnancy due dates [9]. Patients are willing to use PHSs, as revealed in a survey of 800 American patients in which 80% of the patients preferred a patient-centered approach as they felt excluded in the management of their data [10]. With PHSs, patients can access their health information and share it with other stakeholders to co-ordinate their care [1]. Practitioners can make better clinical decisions based on instantaneous access to data in PHSs [11]. In patient-centered health care environments, the value for patients is increased, health care transaction costs are decreased, patients manage interactions through the appropriate release of their own data, and all health care stakeholders will be encouraged to collaborate with patients and other stakeholders to achieve their goals [1].

Technically, PHSs can be deployed using centralized databases (eg, Health Bank [12], Microsoft HealthVault [3], and PittPHR [13]), distributed ledger technology (DLT; eg, Mint Health [14] and Medicalchain [11]), and more flexible peer-to-peer (P2P) technology (eg, OnePatient [15] and doc.ai [7]).

The detrimental effects of centralized health information technology solutions controlled by economic actors are well-known [16], for example, reluctance to innovate or the creation of data silos [16]. DLT-based PHSs, such as MedRec, which is under development at the Massachusetts Institute of Technology [17], are currently spurring the P2P and decentralization push in the health care domain. However, DLT is a specialized P2P technology that does not align well with the needs of the health care sector and the sensitivity of health information. For instance, DLT systems consume excessive computation and communication resources by requiring redundant computations to ensure a consistent state of the ledger across the network, which makes the logged transactions available to all nodes participating in the network, and they have slow processing speeds because multiple parties have to independently verify transactions and arrive at an agreement [18]. The mismatch between DLT and the needs of the health care sector has a simple cause: DLT was primarily designed as a backbone for cryptocurrencies that require one global consistent record of transactions and can thrive even in environments where trusted counterparties do not exist and might even be malicious [19]. Accordingly, DLT is a P2P technology that is too rigid for the health care context, where it is sufficient for all parties involved in the care of a patient to have a consistent view of a patient's health status and existing trust relationships between parties (eg, the patient-physician relationship) can be leveraged. In this study, we take an information security perspective and contribute to the emergence of PHSs that come with the benefits promised by DLT PHSs, such as decentralization, patient empowerment, and interoperable health systems [18], but are implemented based on less rigid and more flexible P2P technology. We refer to such systems as P2P PHSs.

P2P PHS architectures can be based on hybrid P2P networks (eg, P2HR [20]), approaches that combine centralized and P2P architectures (eg, P2P PHR [6] or the e-toile framework in Switzerland [21]), and highly decentralized networks (eg, P2P-integrating health care enterprise [P2P IHE; 22]). Other examples of P2P PHSs, which are disease-centric, are decentralized systems for Bluetooth-based SARS-CoV-2 (or COVID-19) contact tracing, for example, Pan-European Privacy-Preserving-Proximity-Tracing (PEPP-PT) in Europe [22], Trace-Together in Singapore [23], and Stoop in Austria [24], which are used to notify people when they are near SARS-CoV-2 carriers.

In P2P PHSs, the trust and identity of individual participants do not need to be assured through technology. P2P PHSs provide PHS functionalities locally (on any patient edge device such as mobile phones, tablets, etc) under the sovereignty of individual device owners. Patients can make their health information directly available to other participants they trust without the need for any centralized or distributed nodes to facilitate the transactions. However, P2P PHSs have unique security issues because patients must manage information security for their health information largely by themselves, and even qualified professional administrators are already challenged by the task [25]. The absence of a central entity to act as a trusted computing base on P2P networks [25,26] has profound adverse consequences in terms of security that need to be addressed to reap the benefits that P2P PHSs promise to offer.

Objectives

P2P PHSs raise challenging information security–related questions: How can reliable data backups be implemented? If credentials are lost or compromised, how can they be replaced or blocked? How well is the system protected against unauthorized access? P2P PHSs that are not DLT-based (eg, OnePatient [15] and P2P PHR [6]) are an emerging phenomenon that will become more relevant in the future as they are aligned well with large-scale efforts to re-decentralize the internet (eg, the Solid project by Tim Berners-Lee [27]) and support patients in taking ownership of their health data [1,10]. Although P2P

PHSs have been under development for over a decade [21], the dedicated literature on P2P PHSs is sparse. To date, previous studies have focused on security, privacy, and end-user features on centralized and DLT-based PHSs [2,28-31] and did not address security engineering specifically for P2P PHSs, which comes with its own challenges due to a different underlying architecture. To address this gap, this study focuses on security engineering for P2P PHSs based on a systematic literature review. We aim to answer the following research question:

Research question: What are the inherent security issues for PHS deployment on P2P networks and how can they be overcome?

Security issues are defined as any action that could be used to disrupt the functionality of the P2P network or enable unauthorized users to access, modify, or delete user data [32,33], specifically, due to threats or vulnerabilities, such as malware, bugs, access control failures, or patients' inadvertent exposure of their data. To answer the research question, we aim to review existing P2P and P2P PHS architectures and their design choices, study existing PHS features, and propose a suitable architecture for PHS deployment on P2P networks. Thereafter, we aim to highlight the causes and consequences of existing security issues in P2P PHSs and evaluate them based on the identified P2P PHSs in the literature. On the basis of these P2P PHS architectures, we propose security measures for secure provision. To overcome the challenges on the path to P2P PHSs, secure safeguards must be put in place to ensure that information is securely transmitted and protected against cyberattacks [1,34]. Information security is essential for P2P PHSs and will, if appropriately implemented and addressed, increase patients' intention to use P2P PHSs [2,30].

Theoretical Background

P2P PHSs and the Need for Information Security

P2P technology for the provision of PHSs can be flexible and inexpensive for users because it uses available devices at the user's end for deployment. The characteristics of P2P systems, such as fault tolerance, security and trust, scalability, availability, self-reconfiguration, and extensibility [35,36], facilitate and suit the provision of PHSs. With millions of users worldwide, P2P systems have shown strength in providing services for sharing resources without the need for a central server, for streaming multimedia content with distributed load balancing, for volunteering of computing resources, and for telephony applications. P2P PHSs, such as OnePatient [15] and P2P PHR [6], leverage the power of P2P networks and mobile technology to store health records locally under the control of device owners, thereby increasing patient empowerment and control and simplifying the implementation of data protection principles [8,37,38]. P2P systems have better scalability because operations can be executed locally and customized for different purposes. Patients can easily manage access to their health records by using a single-hop connection (eg, Wi-Fi Direct) with other trusted parties (eg, a physician) without requiring a wireless access point or another intermediary communication network.

Factors that impact the security of centralized PHSs are the database size, the large number of potentially affected users, and the confidentiality of the stored data. The health care sector experiences more data breaches than any other sector [39]. A breach barometer in the United States reported 503 breaches for health data in 2018, affecting over 15 million patients [40]. Similarly, the almost immutable nature of data storage in blockchains makes it nearly impossible for users to erase their stored (metadata) information, which conflicts with the European General Data Protection Regulation (GDPR) [41]. Table 1 outlines the main advantages and disadvantages of P2P PHSs.

For patients to benefit from the advantages of P2P PHSs, the network needs to be robust and fault-tolerant. Information security is paramount because of the high sensitivity of medical data [30,42]. Therefore, a pertinent question is how to make P2P PHSs resilient to attacks. P2P systems communicate over the internet; therefore, they inherit the same security issues as any other networked application on the internet. The P2P architecture poses significant security issues such as index poisoning attacks [43], Sybil attacks [44], chatty peer attacks [45], or distributed denial-of-service (DDoS) attacks [46].

Table 1. Security advantages and disadvantages of peer-to-peer patient-centered health care information systems (P2P PHSs).

Dimension	Advantages	Disadvantages
Privacy management	Patients technically govern data. Patients can define access rights to their own PHSs.	Inconsistent views in the network allow attackers (and super users) to cheat and remain undetected.
Federated medical data	Patients keep their medical data and software on their own devices. Patients can determine the desired redundancy for their data by backing up at their end.	Patients may lose access when the device is lost, and no backup system is used by the patient.
Security	No central attack profiles.	Specific security issues other than general networked application attacks are introduced and slow deployment of security patches by users results in insecure P2P systems.
Offline capability	Data are available without a network connection, which improves infrastructure resilience. Disrupted internet connections will not stop data access.	Maintenance effort for storing large amounts of data offline can be high.
Stakeholder interaction management	All health care stakeholders requiring access to patient data have to interact with patients to achieve their goals.	Increased access control requirements for patients are hard to satisfy with current health care processes and systems due to bureaucracy and diverse levels of digitalization.

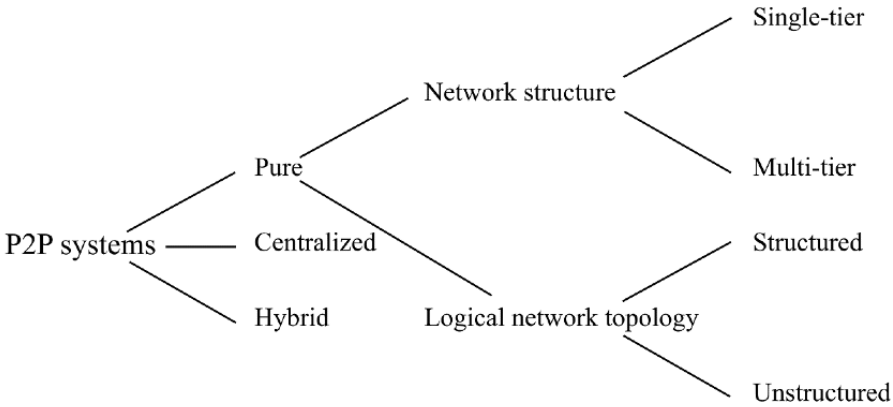
Moreover, P2P systems increase the attack surface owing to 3 disadvantages [26,47]: (1) increased chances of exposing network traffic patterns to attackers; even with encryption, the metadata can still reveal information to external attackers; (2) an inconsistent view of the network (due to a lack of global information), which affects integrity by allowing attackers to cheat and remain undetected; and (3) increased vulnerability to internal attackers due to the absence of a central entity to detect malicious insiders and govern software and security updates.

P2P and PHS Networks

Origins

The concept of P2P was introduced in 1969 in the first Request for Comments of the Internet Engineering Task Force; Request for Comments-1 denotes a *host-to-host connection* [48]. UseNet [49], a distributed messaging system, is often described as the first true implementation of a P2P network and was established in 1979. UseNet looks like a client server model from users' point of view. However, servers communicate with each other based on the concept of P2P and share content over the entire group of UseNet servers without a central entity. With the surge in popularity of P2P networks, the music and file-sharing P2P application Napster [50] was introduced in 1999, which exhibited some approaches to P2P networks known today. Later, well-known and popular P2P systems emerged, such as Gnutella, eDonkey, and BitTorrent. Within the last 2 decades, the first health information systems were deployed on P2P networks—for example, the e-toile P2P PHS framework aimed at connecting all health care stakeholders in Geneva, Switzerland [21,51]; P2HR [20]; or the PEPP-PT COVID-19 contact tracing system in Europe [22]. The features distinguishing P2P systems from centralized systems are peer and resource discovery [35]. Since there are no servers, peers (eg, patients, practitioners, or PHS providers) must rely on techniques, such as indexing and routing tables [52], to locate other peers in the network (Figure 1).

Figure 1. Peer-to-peer (P2P) architectures. Some P2P systems are supported by centralized servers, other P2P systems attempt to decentralize as far as possible. Between these two extremes, hybrid systems benefit from the properties of both.

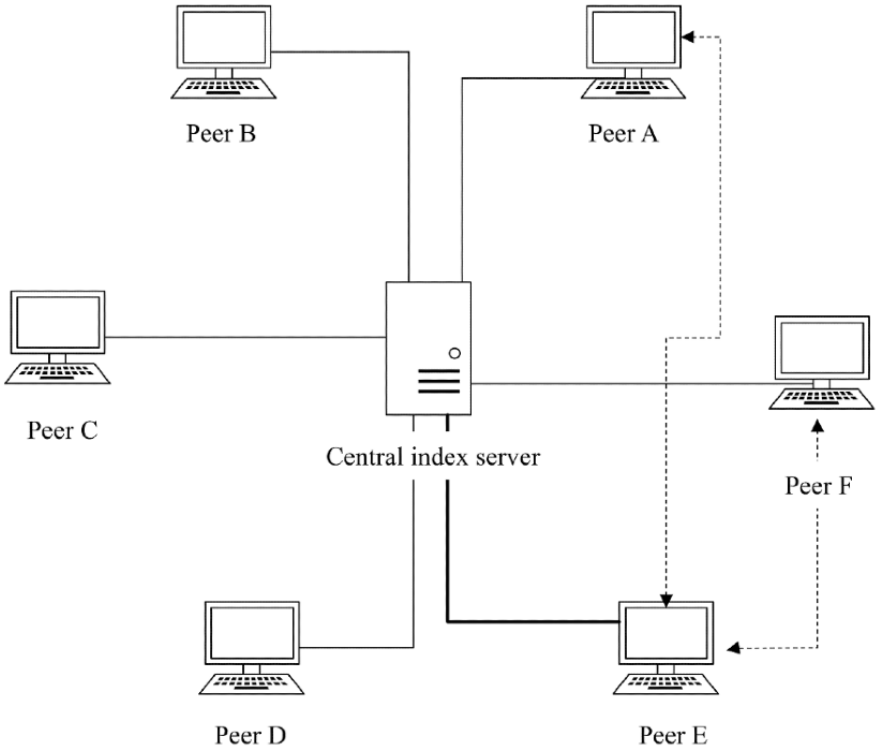


A P2P network, or system, is a type of computer network that exhibits decentralized control, autonomy, virtualization, and sharing of computing resources [47,50]. Peers participating in the network form a P2P network of nodes and are equally privileged. The network is self-organizing. Peers in the network make their resources directly available to other peers without the need for a central entity to facilitate or co-ordinate transactions [35]—for example, patients can directly exchange information with practitioners over their P2P PHSs. Peers in a P2P network can share and download resources. This is in direct contrast to traditional client-server networks in which resource-sharing and downloading are performed by distinct actors (eg, in PHRs such as Google Health or Microsoft Health Vault).

Centralized

Centralized P2P PHS (eg, P2P PHR [6] and e-toile framework [21]), and other centralized P2P systems (Napster, SETI@Home, and BOINC [35,50]) combine the features from client-server and decentralized architectures. One or more central servers are used to manage administration, transaction, registration, or resource discovery. To abide by data protection regulations, such as the US Federal Health Insurance Portability and Accountability Act (HIPAA) [6] or the GDPR [34,41], and related regulations, health or personal information should be stored separately from centrally managed operational data (eg, status and metadata of transactions as in P2P PHR [6] or the list of interoperable PHS providers and health care professionals and their access rights in the e-toile framework [21]). In the case of contact tracing systems such as PEPP-PT COVID-19 [22], the central server may be operated by a government or trusted entity to generate identities and contact graphs. In centralized P2P PHSs, the resources are indexed by the central server (Figure 2). Although a client-server approach is used for resource discovery, the actual communication that facilitates resource transmission is decentralized [53].

Figure 2. The centralized peer-to-peer (P2P) system. A peer E sends a message to the central server asking for the desired resource, the server runs a lookup and determines the peers that contain the queried resource and then sends back the result to the requesting peer E. Once peer E obtained the list (which consists of peer A and peer F), it establishes a direct connection to the peers.



In centralized P2P PHSs, data protection and security measures based on regulations such as HIPAA [6] or GDPR [41] can be enforced and implemented but PHSs may inherit issues from centralized systems [35], such as vulnerability to insider attacks and function creep by the entity running the server; reduced tolerance to avoid single points of failure; and issues with scalability and robustness. Central servers also become more likely to cause a bottleneck when the number of peers increases.

Decentralized

In decentralized P2P systems, peers have equal rights and responsibilities [35,54]. This can be seen in agent-based co-ordination frameworks proposed for the exchange of electronic health records between different providers (eg, P2P IHE [6,51]) or other P2P systems (eg, BitTorrent, Gnutella, Freenet, Chord, and PAST [35,50]). Each peer shares data that may only be relevant to queries of other peers. A decentralized P2P design is a user-based infrastructure because it requires no specific additional infrastructure and depends solely on the participating users to share resources (bandwidth and storage) [26]. In a decentralized P2P system architecture, 2 further dimensions are important [35]: the *network structure* and *logical network topology* (overlay network).

The *network structure* of a P2P network can be *single-tier* or *multitier*. In a *single-tier* network (eg, Gnutella, Freenet, and PAST [35,50]), loads and functionalities are equally distributed among the nodes participating in the network. In contrast, the *multitier* network has a routing structure with hierarchical layers. An example of a P2P protocol in this category includes the Super-peer Architecture and Crescendo System [35].

The *logical network topology* can be *structured* or *unstructured*. In *unstructured* P2P networks (eg, FreeNet, Gnutella, and KaZaA [50]), which exhibit a mesh topology [26], each peer maintains the list of its neighbors to which it may forward queries. Hence, in most cases, a peer must search a large fraction of the network when looking for a desired resource in the network, as there is no precise mapping between the identifiers of resources and peers [55]. Messages are continuously propagated by neighbors in the network [26], which affects the reliability of message delivery when the network is congested. This type of P2P system can be unsuitable for PHS deployment, especially in emergency situations where a patient's medical history (located with another remote peer) is urgently needed for medical care.

To address these problems, *structured* P2P PHSs such as P2P IHE [51] and other *structured* P2P systems (eg, Chord, Kademlia, Pastry, and CAN [35]) have emerged. In *structured* P2P systems, a mapping between peers and data exists, data placement is under the control of Distributed Hash Tables (DHTs), and each peer has to maintain routing tables. A DHT is a hash table containing a key-value lookup function, and the entire index is equally distributed among participating peers [55]. The key-value store represents only the metadata of the participating peers, for example, the mapping (id, ptr) indicates that a resource with identifier *id* is located at a peer pointed to by *ptr*. The general idea of structured P2P networks is to minimize the number of peer lookups (eg, by adopting a key-based routing strategy) to identify and locate a desired resource in the network [35]. The cost of maintaining the structured topology is high when participants arbitrarily join and leave the network.

The overall issue of decentralized P2P systems is the slow search for peers offering the desired resources in the network [35], and freedom to join or leave the network affects availability [20,56]. However, these systems do not have single points of failure and benefit from other features, such as scalability and robustness to operational errors. The lack of centralized control is a major factor contributing to routing difficulties: routing becomes more complicated with more diverse participating nodes [57], when massive peer churn is present [58] and when there is a dependence on nodes that could be malicious [59]. To remedy this, a shared memory in a distributed tuple space architecture [60], as used in the P2P PHS agent-based co-ordination framework P2P IHE [51], can be leveraged. In such an architecture, a distributed network of tuple centers is used as a co-ordination framework to facilitate interactions between various PHS providers and other health care stakeholders [51].

Hybrid

P2HR [20] is an example of a hybrid P2P PHS. Other P2P systems (eg, BestPeer [35], *BestPeer++* [61], or BitTorrent [62]) eventually relied on this topology. Hybrid P2P architectures were introduced to address the challenges of centralized servers in P2P networks and the time required for resource discovery in decentralized P2P networks [35,54]. They combine the advantages of both architectures [50], such as reliable resource discovery and scalability. Although there are no servers in hybrid P2P systems, peer nodes that have more resources in terms of storage, computation power, network connectivity, stability, and uptime can fulfill the role of servers and assist *common peers* with resource discovery. These nodes are referred to as *super peers*. In hybrid P2P systems, resource discovery can be performed by querying the *super peer* (in a centralized manner) or using decentralized search techniques [63]. *Common peers* form the lower layer, while *super peers* form the upper layer.

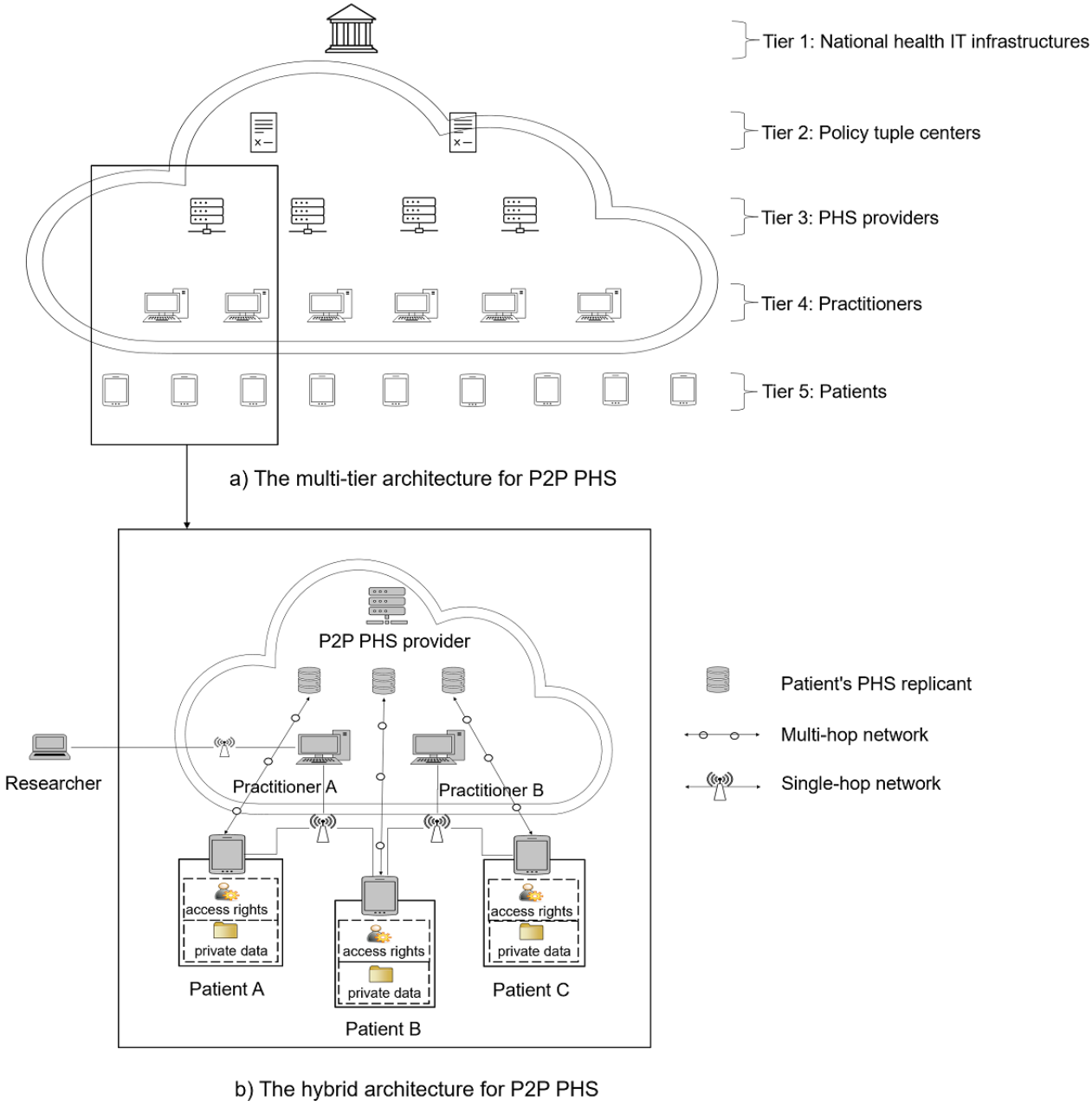
Although *super peers* share some similar properties with servers in a centralized P2P network, they are different [35]: (1) a *super peer* only acts as a manager for its subset of peers in the network—it is not as powerful as a server in centralized P2P networks that oversees the entire network. For PHSs, dividing patients into groups (eg, per hospital) ensures that patients' data are only shared with users that require them [64]; (2) a *super peer* also participates and acts as a *common peer* and facilitates the same operations, such as resource-sharing and downloading. As an analogy, the relationship of *super peers* with *common peers* is similar to interactions between entities in human society: for instance, in a hospital, physicians keep more knowledge and connections with their patients than other personnel. As such, patients with health issues are expected to ask for help from physicians, as there is a higher probability that they are able to handle the problem.

Super peers can act as *federated* authorities whereby participating users can affiliate themselves with provider nodes based on extant trust relationships (eg, friendship or treatment relationships). Provider nodes are largely independent of each other; hence, there is a federation of provider nodes. Each provider is responsible for its common peers; however, individual provider nodes can collaborate to provide services. The placement of *super peers* in a privileged position enhances the availability of resources, operations, computations, and performance; however, this also raises issues regarding trust, privacy, and integrity as *super peers* regulate services. The absence of a *super peer* in the network may affect operations in the network, thereby reducing the fault tolerance of the P2P network. In terms of security, nodes operated by providers are central points of attack (at least for the common peers served by a particular super peer). As super peers manage subsets of peers in the network, they are more attractive targets for attacks. "The main vulnerability of federated systems are such assumptions that federated service providers (e.g., super-peers) will largely act honestly" [26].

P2P PHS Architecture

On the basis of the discussion of the different forms of P2P PHS architectures in the previous section, the combination of multitier structure and hybrid P2P architecture appears to be most appropriate for P2P PHSs; therefore, we propose an architecture with the following abilities (Figure 3): (1) enforcement of data protection requirements similar to that of HIPAA and semantic compliance through *super peers* as central index servers; (2) registration and identity verification; (3) higher scalability and availability of resources and lack of single points of failure; (4) association of patients (tier 5, Figure 3) with their respective PHS providers (tier 3, Figure 3) and practitioners (tier 4, Figure 3); and (5) faster PHS updates with security patches through the *super peer* networks. The P2P PHS network is an overlay of the modeled hierarchical relationships between the tuple center and PHS providers, PHS providers and practitioners, and practitioners and patients.

Figure 3. Proposed peer-to-peer (P2P) high-level architecture for patient-centered health care information system (PHS). An aggregate relationship exists between the practitioners and the patients. The patients control the access to their health data, and other entities require patient permission to access a patient's medical data, for example, by using tokens as currently being implemented in the MedicalChain PHS project [11].



Large health care IT organizations (eg, the German Healthcare Technology Infrastructure; HTI [2,65]) are represented at the top of the hierarchy in the architecture to facilitate certification of various PHS providers (tier 1, Figure 3). They define and enforce the implementation of various data regulations, representation standards, and ontologies (eg, Health Level Seven and Fast Health care Interoperability Resources [6]) to share heterogeneous medical records across PHS networks. In the second tier, a distributed public network of tuple centers (eg, certified through a national health agency) is provided by trusted third parties (tier 2, Figure 3). Agent-based systems (as in centralized P2P PHSs [51]) can be used across P2P networks with the tuple centers' action-reaction rules for communication events [51]. Agent co-ordination models can handle services for data semantics and peer lookup services while serving as mediums for data sharing between P2P PHS providers, but the actual inter-PHS communications are performed in a P2P manner. P2P PHS providers can subscribe to any certified tuple center. Communication of a PHS provider is limited to communication with other subscribers to the PHS provider's tuple center subscriptions.

PHSs can be provided by any party. In our scenario, we exemplify hospitals (*hyper peers*—managers of super peers and other peers in the network) as PHS providers. The *hyper peers* relay requests and responses among all subpeers across multihop networks. Each *hyper peer* has its own separate private cloud server, which stores a digital and secure copy of patient health records (Figure 3). These records are a replica of the data available on the patient's local storage but are only made available in the *hyper peer's* private cloud if a patient subscribed

to the corresponding additional PHS features (eg, for data backup, ease of remote data sharing, or emergency access). Accessibility and availability traits of the stored *common peers'* data on the private cloud are in the control of patients through their local PHS client software. This topology can have 2 issues: (1) similar records of patients are stored locally on their mobile devices and the cloud, which appears redundant, but this redundancy curtails connectivity pitfalls while preserving P2P PHS features in terms of offline capability, and (2) the cloud storage can become inaccessible when the local patient PHS device is lost when the device is used as the source of patient identity verification and access authorization for cloud storage.

Each *hyper peer* has multiple health practitioners in the network, which maintain patients' public identities (under the control of DHT [55,66]) for lookup functionality and ease of data access; therefore, a patient (*common peer*) can be associated with multiple practitioners from various *hyper peers* (practitioner A, B, C, etc). In such cases, these *hyper peers* can communicate via tuple centers. This way patient data stored on a cloud of hospital B can be accessed by practitioners in hospitals A or C for diagnosis or treatment, given that the patient grants access rights. Each *common peer* on the network (corresponding to a patient) is modeled on the local PHS and on the *hyper peer's* private cloud server. *Common peer*s can grant access to their health records to any party through single-hop radio communication (without involving a third party in the communication, eg, Wi-Fi direct) or multihop network communications via the cloud storage of the *hyper peers* [65]. Other parties, such as researchers looking for data for research purposes, can obtain read-permissions for patient records by interacting with the practitioner via the hospitals' private network, which forwards permission requests to patients. However, only aggregated results (anonymized) are returned to the researcher. Moreover, wearable mobile devices and biotechnologies that provide biometric or psychometric data can also be directly connected to a patient's P2P PHS.

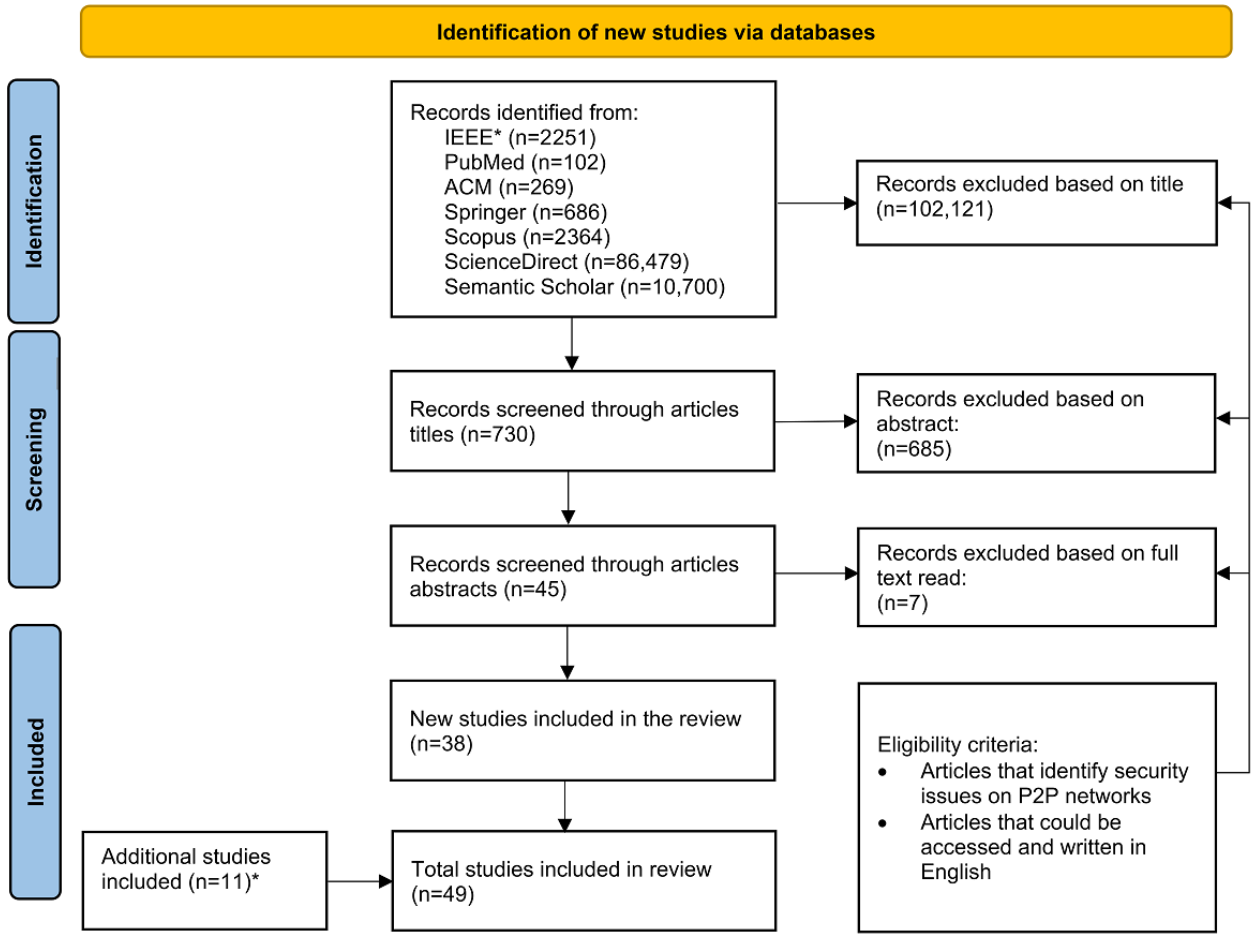
Methods

Literature Search

We conducted a systematic literature review (Figure 4) following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines [67,68] and used thematic analysis to guide the data analysis process [69]. The systematic literature search in this study was conducted using specialized academic search engines (IEEE Digital Library, PubMed, Science Direct, ACM Digital Library, Scopus, and Semantic Scholar; see Multimedia Appendix 1 for further details). The search was conducted on articles published between 2008 and 2020. The study selection was organized into the following phases.

1. The search string was derived by breaking down the research question into different facets, where their alternative definitions and acronyms are included and combined using the logical operators "OR" or "AND" [68]. The search string "(*P2P* OR *Peer-to-Peer*) AND (*vulnerabilities* OR *vulnerability* OR *threats* OR *threat*)" was applied to the title and abstract and adapted to the specific syntax of the used search engines.
2. Eligibility criteria: we included all articles that could be accessed, were written in English, were published in academic outlets, and identified inherent security issues for PHS deployment on P2P networks, as suggested for thematic analysis [69].
3. Abstracts of the filtered articles were further analyzed by the authors to remove irrelevant articles based on eligibility criteria and other false-positive results.
4. Articles were grouped and duplicates were removed.
5. The remaining articles were read in full text and analyzed by the authors (assisted by Atlas.ti software [70] to manage codes and themes for thematic analysis [69]) to include only relevant studies based on the eligibility criteria defined in step 2.

Figure 4. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. P2P: peer-to-peer.



Identified Articles

Initially, 102,851 articles were identified using the search string. The filtered articles were screened based on their titles using the same search strings. A total of 99.29% (102,121/102,851) false-positive results were removed. Further examination of the abstracts of the remaining 0.71% (730/102,121) articles resulted in the exclusion of 0.67% (685/102,121) articles. The main reason for exclusion in this step was a lack of thematic fit with our study (eg, a focus on P2P currency exchange or lending platforms or security issues for largely unrelated technologies such as robotics). We analyzed the full text of the remaining 0.04% (45/102,121) articles, and 0.01% (7/102,121) further articles were excluded. We complemented the result set with 0.01% (11/102,121) additional articles that met the eligibility criteria but not the inclusion criteria (eg, published before 2008). Ultimately, 0.05% (49/102,121) articles remained.

Thematic Analysis

Data analysis was guided by thematic analysis [69] to identify the relevant themes in the identified articles. The initial coding was performed by the first author and refined and finalized in group discussions with the other authors. The themes (codes) were identified using the key security goals (theory-driven) from the CIA (ie, confidentiality, integrity, and availability) triad as organizing codes for data analysis (assisted by Atlas.ti software [70] to manage codes and themes for the thematic analysis). *Confidentiality* entails that unauthorized actors cannot access information during transmission, processing, or in storage. *Integrity* requires that the information not be modified unintentionally or without authorization. *Availability* means that the system is accessible to the user when needed. For each of the codes identified, we looked at the impact of the security issues associated with the codes to examine their impact on P2P PHS (eg, potential for unauthorized access). We then investigated and rated the consequences of potential exploits of P2P-PHS security issues based on the Common Vulnerability Scoring System (CVSS; see Multimedia Appendix 2 for further details).

The systematic literature review revealed 8 main P2P security issues (list of themes) extracted through data analysis and 7 factors promoting them. Table 2 shows the summary—generated codebook—of the security themes identified along with their sources and exemplary codes used to derive the themes during the thematic analysis process.

Table 2. Overview of peer-to-peer security themes identified[a].

Combined themes, second-order themes, and first-order themes		Study
Pollution		[71-81]
Metadata pollution	• Changing original file name or extension • Replacing the file with a misleading one	
Index pollution	• Claims ownership of wanted but bogus content • Sharing of the content record via the index	
Content pollution	• Modifying the file content • Replacing the file with an incorrect one	
Malware		[78,81-91]
Virus	• Infection of the system • Appears to be part of legitimate programs	
Spyware or ransomware	• Spying or stealing user data • Encrypts any kinds of files and data	
Worm	• Infection of P2P[b] routing table • Appears independent of existing programs	
Social engineering		[35,80,82,92-95]
Baiting	• Tricks user to divulge sensitive information • Relies on human error or mistakes	
Phishing	• Scam via email or SMS text messages • Trick into divulging sensitive information	
Poisoning the network		[35,43,45,47,56,71,73,77,81,89,95-102]
Index poisoning	• Sharing of bogus contents via indexing table • Affects network quality of service	
Routing table poisoning	• Sharing of bogus contents via routing table • Prevents from finding correct resources	
Sybil		[26,52,54,56,63,72,76,81,84,92,103-112]
Faking identity	• Faking multiple identities for a single user • Affects the redundancy property of P2P systems	
51% attack	• Outvoting of honest nodes in the network • Cheating without being detected	
Eclipse		[47,54,56,72,77,79,81,92,93,105-108,113-116]
Large man-in-the-middle	• Separating the network into several portions • Acts as gateway and disrupts message flow	
DDoS[c]		[43,45,72,76,77,80,81,84,88,92,94,95,97,98,100,102,105,110,117-119]
Flooding	• Invalid packets flood the network • Impedes delivery of normal packets	
TCP-DDoS[d]	• Connection overload with full TCP-requests • Denies connections from legitimate requests	
P2P traffic blockade		[46,100,120-122]
Port number blockade	• Blocking of P2P network traffic • Imposes bandwidth limits with P2P networks	

[a]The first- and second-order themes are only examples and not exhaustively listed.

[b]P2P: peer-to-peer.

[c]DDoS: distributed denial-of-service.

[d]TCP-DDoS: transmission control protocol–distributed denial-of-service.

Results

Factors Promoting Security Issues in P2P Networks

To use a P2P network for resource-sharing, multimedia-streaming, distributed-computing, or telephony applications, users install a P2P application on their device and permit the application to access and use device resources such as cameras, microphones, or device storage. In P2P operation, the P2P client application reads files from the user's disc during the uploads and writes to the user's disc during download. During this operation, personal or sensitive information can be transmitted to the network.

Inadvertent Sensitive Information Disclosure

It is often not necessary that users' confidential or personal documents be exposed by worms or viruses, as many users inadvertently expose these documents [123]. For example, a node may request data X from the user, and the user sends back the entire folder where data X is located. The user may end up exposing all of their sensitive information for the following reasons: (1) a user does not appropriately select or share the requested data, (2) the interface design of the P2P application confuses the user, and (3) the requester offers a huge incentive to share. In 2012, an automated personal health information tool was used to crawl different P2P networks (FastTrack, Gnutella, and eD2K) to analyze Canadians' personal health information and personally identifiable information in the exchanged text files [83]. Out of the 3924 P2P files with unknown content, 1.45% (57/3924) of files were flagged as personally identifiable information. Manual analysis of the 57 files revealed that 19% (11/57) contained health information about an identifiable individual, that is, inadvertently disclosed health information.

In 2019, a survey identified human errors, such as sending personal information to unintended email recipients or releasing personal information by accident, as the largest source of data breaches in the health sector [39]. Similarly, several peers were found to be inadvertently sharing their financial, email, and web cache data in a study on the KaZaA P2P network [124]. In addition, some P2P users share their personal information intentionally to increase the number of files shared on the network to meet the participation requirements of some P2P systems [85].

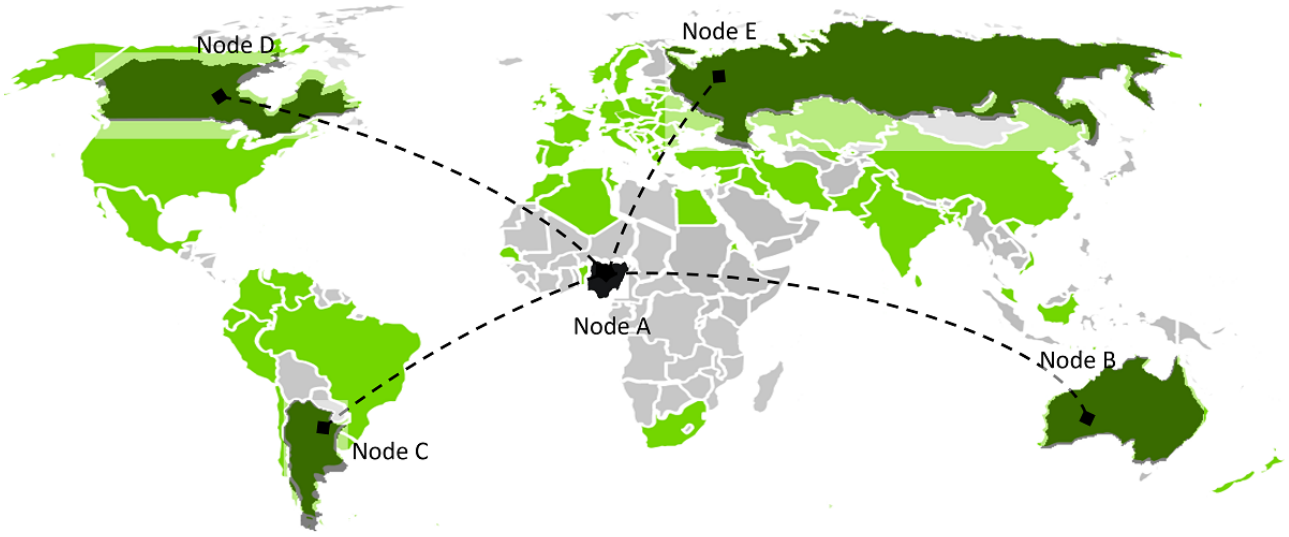
Set-and-Forget

P2P clients tend to be set-and-forget applications that run in the background [85,123,125]. This means that the user is not cautiously tracking the activities of the P2P client, which increases the opportunity for abuse.

No Borders

Geography is largely irrelevant in P2P networks [85], and no region is safer than the other. A computer in Australia or Argentina becomes part of the same network as a computer in Nigeria (Figure 5). In open P2P networks, files can undoubtedly migrate globally, and threats can come from any region of the globe. Hence, the heterogeneity and geographically dispersed nature of P2P networks can be a problematic factor affecting security, quality of service guarantees, and scalability. However, studies have shown that P2P networks converge to a certain degree of geographical clustering [85,126]. Users may choose to download and share content from their region to have lower network use and latency than when downloading or sharing content outside their region.

Figure 5. Geography example of a peer-to-peer (P2P) network.



Growing Use and Network Heterogeneity

As a P2P network grows, an increasing number of leaks of confidential files will occur in the network. In 2017, nearly 27 million P2P users downloaded and shared files on P2P networks daily, which is 17 million more users than in 2006 [127,128]. Moreover, P2P networks are heterogeneous and fast-moving; hence, users may not be able to keep track of security issues and developers may neglect them [85].

No Content Verification

Conventional P2P networks have no trust mechanism to assist users in deciding whether to share or download content in the network. Similarly, they have no central authority responsible for verifying the authenticity of the resources shared by users [80]. Hence, there is no guarantee that users are sharing the content they promise. This makes it easier for an attacker to spread malware across a P2P network, for instance, to conduct fraudulent activities or pollution attacks [72].

Digital Winds Spreading Files

Typically, P2P networks create file indexes using the names of the files and the associated metadata [123]. This constitutes a security issue, as it allows anybody to easily discover files in P2P networks. For example, an opportunistic search with key terms related to the top 10 publicly traded health care firms in the United States revealed 20,000 patient records, 4 patients with acquired immune deficiency syndrome (AIDS), 201 patients with a mental diagnosis, and 326 patients with cancer [125]. The approaches that some P2P clients use to create and manage file names have serious implications in exposing users' private and confidential information. This can be a problematic factor regarding security because users' sensitive files can be easily discovered owing to poor P2P client design.

Snooping Nodes

This factor enables attackers to leverage the open nature of P2P networks [100]. The long routing paths across several nodes create a loophole for malicious activity [94]. Peers in a privileged position in the network (eg, *super peers*) are able to see the communication of other *common peers* in the network. For example, decentralized P2P systems such as Gnutella [35] have no central servers or auxiliary mechanisms to co-ordinate communication among users, but when a new user connects to the Gnutella network, it chooses a node as its permanent entry point [115]. Thus, high-speed nodes are inadvertently placed in the central part of the topology and can observe the communication of nodes in their local subgraph. Moreover, communication in P2P networks stops being anonymous as soon as the source node establishes a direct connection to a destination node to download files [35]. The IP addresses of both nodes are exposed to each other, which creates another opportunity for abuse. Once the identity of the peer is revealed, further attacks can be carried out [96].

Identified Security Issues and Their Impact on P2P PHSs

Pollution

Pollution is a form of attack in which an attacker modifies the original content (through mixing or substituting) so that it has no use or is of low quality [72,79,81]. The polluted content appears to be legitimate content (eg, by having a similar size, format, and title) to trick users to download it. However, the altered content may be malicious, fake, or corrupt. This affects the network's quality of service (especially in file, voice, and video-based P2P streaming systems [72,73,75,79,80]), overall system energy consumption [74], content availability [78], and data integrity [72]. Pollution is an easy and fast way to disseminate worms or viruses from one to many peers in the network. Therefore, pollution can have an exponential impact on the security of the entire network [72]. The pollution attack was first discovered in 2005, where a crawler was used to retrieve super peers in the KaZaA P2P network [73]. Analysis of the contents collected by the crawler revealed that over 50% of welcome copies (ie, introductory files for a collection of files) for musical files in the KaZaA network were polluted [73]. Pollution is a serious attack on P2P networks, even in a scenario with only one polluter [72,75]. The impact grows when the number of polluters or peers attempting a request increases [75]. As a result, peers often require multiple times the network bandwidth they need in a network free from pollution [75]. Furthermore, the attack is persistent. Even if the polluted contents are identified and blocked by the network, the polluters may remain alive in the network by disguising their identities and can keep polluting the network.

Pollution is categorized based on the attackers' strategy: (1) metadata pollution, where a file extension or name is modified and replaced with a misleading one; (2) content pollution, where the file content is changed; and (3) index pollution, where an attacker claims ownership of an unindexed bogus file and uploads its record (IP address, port number, etc) to the entities (eg, super peers on hybrid P2P) that maintain such records for distribution [73,77]. In most cases, the polluters also attack legitimate peers' reputations or boost their own reputation through whitewashing attacks [75,76]. Content pollution is the most popular and common attack in P2P streaming systems [74]; it was detected in 50%-80% of files in KaZaA and about 50% of popular files in eDonkey [73,74]. Pollution is not necessarily caused by malicious users; P2P systems are notorious for illegally sharing and disseminating copyrighted content, and content is often polluted by copyright owners as a countermeasure to protect their rights when legal actions fail [71,72]. To facilitate the protection of copyright claims, some P2P system providers even weaken protection from pollution attacks in their network [73], although this affects the confidence of users in such systems [72,73].

Impact of Pollution Attacks on P2P PHSs

Successful pollution attacks on P2P PHSs can be devastating because of the higher integrity and availability requirements of medical data than data shared in other P2P systems. The consequences of its exploitation could be between low and high, depending on the level of access gained; pollution attacks often serve as a gateway to identify vulnerabilities (eg, unverified inputs that can be used for SQL injection attacks [129]) and mount further attacks (eg, ransomware attacks). For example, in 2020, a patient in need of emergency care due to an aneurysm died in Germany during a ransomware attack in a hospital. The ransomware attack caused a network outage that disrupted

emergency services, and the patient was sent to a health care facility approximately 20 miles away [130]. This diversion delayed the treatment of the patient by an hour and she died [130]. The openness of P2P systems allows polluters to easily join and leave the network [20,56]; however, identity verification (eg, via insurance, job contract, token, etc) and multifactor authentication concepts for P2P PHSs could create an additional layer to reduce the vulnerability of the network. Patients or practitioners polluting a P2P PHS through their legitimate accounts can easily be traced; however, in some situations, a double-faced user (legitimate but malicious) could leverage open-source hacking tools such as Burp Suite [78] to, for instance, alter an http request payload with an anonymous ID, add polluted content, and forward it to the content distribution network of a hospital to harm the network.

Malware

Malware refers to a wide range of attacks that compromise a system without the knowledge of the system owner [84,90]. P2P networks present a greater risk for receiving malware; for example, only 3 strains of malware infected over 68% of compressed and archived files on the Gnutella network [84]. In the first 3 quarters of 2019, 7.2 billion malware attacks were reported globally [91]. In P2P networks, malware is predominantly used to create botnets by leveraging worms [84,89,90].

A botnet is a network of infected nodes that are usually compromised by worms or viruses. Individual bots in the botnet only use a small portion of the infected resource to remain concealed and create only barely noticeable traffic to share data from the compromised computers with the target [88,89]. The bots are controlled by an attacker (botmaster) through command-and-control servers [89].

A worm is independent and neither requires a host application [84,87,92] nor human intervention [82] to propagate and replicate itself over a network. Worms can result in a high fallout in combination with other vulnerabilities and propagate themselves over email attachments, web server infections, file downloads (counterfeit worms), or other legitimate network activities (silent worms) [78,81,82,84,87]. Passive (counterfeit and silent worms) and active worms are 2 broader categories of P2P worms; they both propagate like a biological virus, but the former waits for victims to infect, while the latter actively searches for new targets [84]. The threats to the amplification of worm-based attacks in a P2P network are high, and the impact grows based on network size, topology degree, or host vulnerability [78]. In contrast to the internet, where worms need to randomly search to identify vulnerable hosts, P2P worms spread rapidly and infect all nodes in the network almost instantaneously [84]. For example, the Antinny (passive and counterfeit) worm that appeared on the Japan-based Winny P2P network led to the disclosure of a large amount of private data: thousands of patient health records, customers' identifiable information, top-secret military information, and documents of a county police investigator, yielding information on major investigations on 1500 individuals [85,86]. Furthermore, in 2001, in less than 14 hours, the Code-Red worm (active) infected over 350,000 systems and caused more than US $1.2 billion in damages in the first 10 days of its circulation [78].

P2P worms are some of the best facilitators of botnet-based attacks and internet worms. P2P networks are, for instance, known for sharing *gray* content, such as pornography and pirated streaming media. This can lead users to incautiously monitor unusual behaviors in the network [78,84,85]. Active P2P worms have different attack strategies: pure random scan (PRS), offline hit-list scan, and web-based scan [78,82,84]. The PRS is a starting point, information gathering stage, and is the most commonly used strategy [78]. PRS is useful when the infected host (bot) possesses no prior vulnerability information of potential targets and randomly selects and mounts attacks on targets to propagate the infection, for instance, using random IP addresses searched from the global internet address space [78,82,84]. The offline hit-list scan is a more powerful strategy: the attacker collects and continuously attacks targets using DNS, network topology, and routing information of P2P systems (eg, using crawler tools [83]) until all the hosts in the hit-list are scanned, and the newly compromised bots attack using the PRS strategy [78,82]. Instead of an offline hit-list, the web-based scan strategy primarily launches attacks on its web-based P2P neighbors, and then the worm disseminates further using PRS through the infected worm hosts [78,82].

Impact of Malware on P2P PHSs

Ransomware constitutes the biggest threat with 151.9 million attacks globally in the first 3 quarters of 2019 [91]. Moreover, ransomware attackers are shifting tactics to target higher-value institutions, such as hospitals [91]. In 2017, a malware was used in the WannaCry ransomware attack, which infected more than 230,000 computers worldwide [131]. In the British National Health Service, WannaCry disrupted scheduled treatments in many hospitals, resulting in total damages of around £92 (US $12.6) million in the United Kingdom [132]. The malware hijacked users' data, encrypted the data, and blackmailed users before decrypting their data [133]. For health data on P2P networks, which have a less controlled infrastructure, ransomware attacks can become easier.

The effect of malware on P2P PHS could be high, although the severity of malware attacks is context-dependent. The effect of malware, such as Antinny [85,86], Anatova [134], or Code-Red [78], on P2P PHSs will be detrimental if it denies patients and physicians access to the PHS, steals patient data, or hijacks and encrypts data for ransom. Structured P2P PHSs, similar to our proposed architecture (Figure 3) or the e-toile framework in Switzerland [21], could be less vulnerable to malware in comparison with unstructured P2P PHSs. This is due to the possibility of using control measures on the index and DHT networks [55,66]. The factors that increase the attack surface include that P2P client applications tend to be *set and forget* [85,123,125] so that they run in the background while the user is not monitoring its activities and that there is no centralized control to detect and prevent attacks in P2P networks. The impact of malware could also escalate beyond the boundary of the P2P network and impede usability features such as emergency access or guardian support. In P2P PHSs, these disruptions can occur on a greater scale than in the example in

the previous section, where a single patient could not be treated in a hospital because of a ransomware attack [130].

Social Engineering Attack

Some P2P clients are being used by users with limited knowledge of computers and information security [80,94,95]. Depending on the nature of the target network, the effect of social engineering attacks—an attack on the users involved in a system [93]—can facilitate exploits of other vulnerabilities. P2P worms such as silent worms (eg, VBS.Gnutella worms [82]) are based on social engineering, disguise themselves, attach to a known file, and wait to compromise victims [93]. Moreover, some P2P systems (eg, Napster and BitTorrent [92]) implement mechanisms in which the users are incentivized to share resources or content to gain greater performance and access to content; therefore, experienced users or attackers can exploit the eagerness and likely incautiousness of new users to deceive them and obtain confidential information, which could be used to conduct malicious attacks. Owing to the *set-and-forget* nature of P2P file-sharing applications [35], users may not realize the breach of confidentiality risks when using them, which increases the chances of abuse.

Impact of Social Engineering on P2P PHSs

Social engineering can affect all types of P2P PHSs, where an attacker can easily leverage the user layer to deceive patients (older adult patients are more vulnerable to this attack than others [135]). In the case of P2P PHSs, the threat impact could be one user at a time, with the probability of escalating and affecting others in the network. Social engineering can be observed as an intelligent information gathering stage for attackers to mount other attacks [129], such as scamming patients to obtain, for instance, access credentials to their P2P PHS accounts. Depending on the attackers' goals, they may modify patients' health records or upload malware to the P2P network to affect patients' lives, health, location, privacy, behaviors, or activities [93] and sabotage the PHS and its providers.

Poisoning the Network

Poisoning can be performed either by *index poisoning* or *by routing table poisoning* [102]. Many P2P systems have a lookup service using indexing or routing table techniques [35,47,95]. A poison attacker can use this to inject invalid information such as bogus resource identifiers or fake IP addresses into the lookup service. An index poisoning attack affects the index of P2P systems [43]. Injecting invalid information in the index or routing table can slow down the query, prevent others from finding the correct resources, or result in a peer wasting time connecting to invalid peers [100,102], which eventually affects the P2P network's quality of service [101]. Some anticopyright infringement organizations use poisoning attacks to prevent the sharing of pirated content on P2P networks [89,99,100]. These attacks are performed by identifying and poisoning the IP addresses of the servers for pirated content or using their IP addresses as evidence to sue the content server or P2P system providers [71].

An index maintains records in a centralized manner (eg, Napster [50], P2P PHR [6], or e-toile framework [21]) and enables users to locate resource owners' IP addresses and port numbers. In *index poisoning attacks*, the attacker aims to compromise indexing peers (peers that participate in the indexing) by adding invalid information into their local indexes by simply sharing the bogus information with the indexing peer [43,81].

A poison attacker can also attack a specific host; for example, if the attacker wants to conduct a DDoS attack on the application server at host 129.13.152.6, the invalid information may include 129.13.152.6 for the IP address and 80 for the port number. Once the indexing peer has been poisoned, another peer can search for a resource and eventually receive invalid information from the poisoned peer and try to download the resource from the victim host. Before downloading the resource, the transmission control protocol (TCP) connection is established with the victim host using invalid information. To download the resource, the requesting peer sends a message to the desired resource. When many peers try to download the resource from the victim host, a TCP-connection DDoS comes into effect [43,97,98].

Structured P2P systems (eg, P2P IHE [51], our proposed PHS architecture [Figure 3], Chord, and Kademlia [35]) are vulnerable to poison attacks [95], although resource discovery is under the control of data structures (eg, DHT). In *routing table poisoning*, the poison attacker exploits the fact that each peer in a DHT-based P2P system maintains the routing tables of its neighbors [47,56,73,77,95,96]. Each entry in the table includes the neighbor's identifier, IP address, and port number. The attacker can deceive participating peers by injecting invalid neighbors into their routing tables. The poisoned peer may choose an invalid neighbor in its routing table and forward its messages. If the routing tables of many peers are poisoned with invalid information and each entry points to the IP address of the victim host, the target receives a flood of messages from the DHT [95]. A further type of content pollution attack is a *combination attack* that combines *index poisoning* and *fake-block* attacks to have a higher impact [45,77]. In this case, poison attackers use an index poisoning attack to include their IDs in the invalid information to be advertised. If the victims establish the connection through the invalid information, they may connect to a poison attacker, so that the attacker can feed the victims with fake fragments and impose more harm on them.

Impact of Poisoning Attacks on P2P PHSs

Centralized P2P PHSs, such as P2P PHR [6] and the e-toile framework [21], could suffer the worst effects of poison attacks because they can cause DDoS or entire network failure and disrupt the services offered by PHSs. For example, in the e-toile framework [21], a list of health care stakeholders and their access rights, data exchange, and authentication is managed by a central index server; poisoning such an index could mean that the data of a patient registered with PHS^X in need of emergency care at a remote hospital that uses PHS^Y could be inaccessible to practitioners. Even if the networks of PHS^X and PHS^Y are not affected, the single point connecting the PHS providers is disrupted. Depending on the urgency of a patient's need for treatment, the need for access to health data, and the longevity of the attack, the patient's health and life could be adversely affected. In some P2P PHSs (eg, P2P PHR [6] or P2HR [20]),

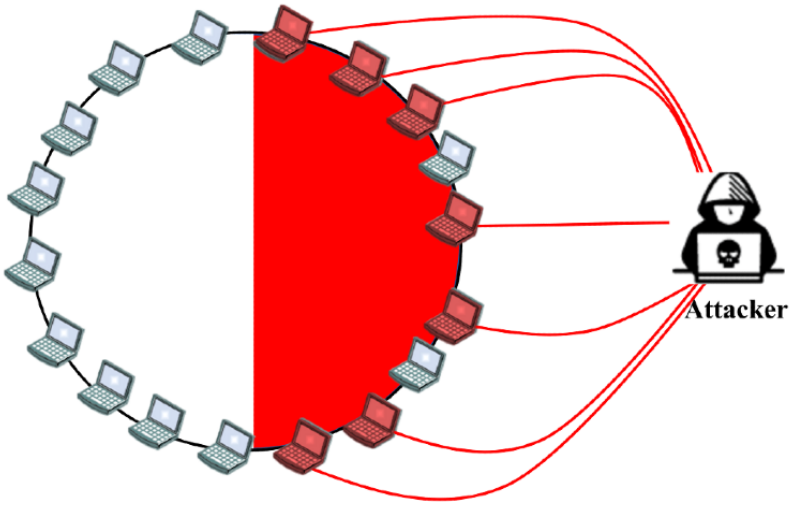
peers' IP addresses are exposed to facilitate health information exchange between different health entities; this makes the attack even easier. For our proposed P2P PHS architecture (Figure 3), there is a federation of PHSs and tuple center providers. Within the context of the previous scenario, access and data exchange will not be impacted if PHS^{Y} is in the same tuple group as PHS^{X}.

Sybil Attack

The name Sybil attack was coined by Microsoft Research in 2002 based on the book Sybil about a patient, named Sybil, diagnosed with dissociative identity disorder [111]. In computer security, Sybils refer to multiple identities of a single user on the same machine; this user can become powerful and control a significant part of the network or use the identities to influence the system behavior [54,56,81,109,110,112]. In DHT-based P2P systems, a user can locally generate multiple *node IDs* for many node instances on the same machine [108]—on the Kad network, a single node can select multiple IDs concurrently [107]. The creation of Sybils is considered the most harmful behavior on a P2P system [54], as it offsets the network's redundancy property [81]. Sybil attacks occur in a P2P network, when the reputation mechanisms are compromised [72], secure authentication mechanisms are not implemented (eg, no proof of identification is required for registration in the P2P session initiation protocol network [106]), or verification of a client's IP address and its maximum number of connections per ID is not implemented (eg, Kad network [98]). Limiting the number of connections per IP address (eg, in eDonkey [84]) does not prevent Sybil attacks because attackers can bypass this by having many virtual IP addresses. It seems that there is no clear and definite solution to prevent Sybil attacks [26]; this is due to the openness and lack of admission control mechanisms in P2P networks.

Sybils are used by attackers to conduct massive and organized attacks on P2P networks [92]. For example, eclipse attacks [54] amplify Sybil attacks through the combination of Sybil and ID assignment or mapping attacks [105], which assigns identifiers near the same portion of the ID space to sufficient Sybil nodes (Figure 6). This enables the attacker to own a deciding power of where in the ID space the new nodes are placed. When the attacker owns more nodes than the benign nodes in the segment, the attacker can control messages in the segment, bias reputation score, create DDoS situations, or force servers to exceed their CPU capacity [26,76,84], which is also known as a gateway attack [92]. In blockchain P2P networks, Sybil attacks are, for instance, used by attackers to outvote the honest nodes in the network [52,63,104], which enables the attacker to cheat without being detected. After a successful Sybil attack, attackers can transmit or discard blocks, effectively block other users from the network, carry out *51% of attacks* to change the order of transactions, prevent transactions from being confirmed, or even reverse transactions that they made, which can lead to double spending [103].

Figure 6. Example of Sybil attack [92]. The attacker placed his malicious nodes on one side of the network segment. Placing many malicious nodes in the network enables the attacker to gain control of the activities of one-half of the network.



Impact of Sybil Attacks on P2P PHSs

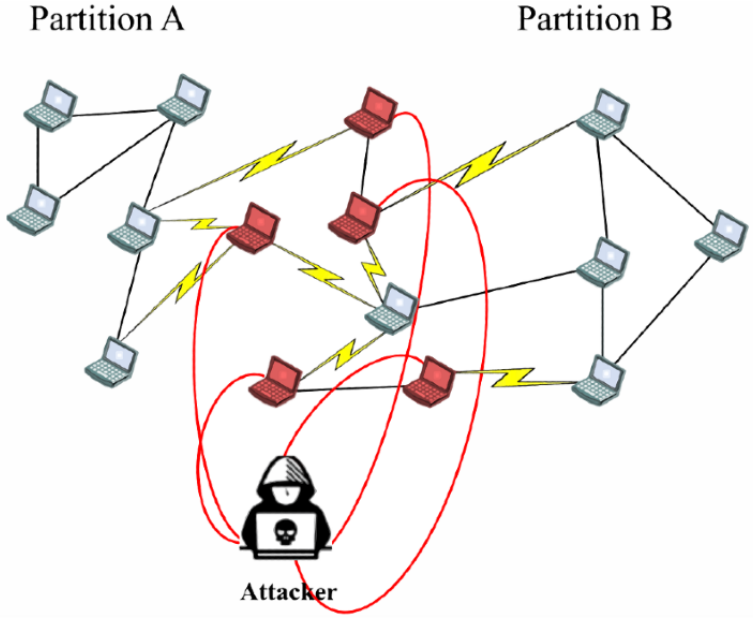
Sybil attacks are helpful for attackers to disguise their identities, access vital information managed in the PHS index service, monitor communications between users, steal patient data, or pollute the entire network to disrupt the entire PHS service operation, which would affect patients' health and life and sabotage the PHS provider's reputation. In our proposed PHS architecture (Figure 3) or the e-toile framework in Switzerland [21], the national health IT agencies are tasked with effectively handling health care stakeholders' registration, authentication, and verification; therefore, freedom to create multiple concurrent IDs on the same system by any malicious user is reduced by design. P2P PHSs, such as P2P IHE [6,51], could be more vulnerable to Sybil attacks due to the difficulty in establishing control mechanisms in a decentralized network. In any case, attackers can leverage Sybil attacks to steal patients' identities (eg, for insurance coverage or blackmail).

Eclipse Attack

An eclipse attack is a large-scale man-in-the-middle (MitM) attack that is commonly executed at the P2P network level [54,92]; routing, sniffing, and traffic analysis attacks are variants [56,79,81,93,105,106,115,116]. An eclipse attack aims to

separate the entire network into 2 or more partitions (Figure 7) by placing malicious nodes in a strategic routing path of the P2P network [105,106,108] to surround benign nodes with malicious neighbors [77]. In most cases, the routing mechanisms are attacked [47]. This is accomplished by adding the attackers' addresses to the neighbor list of the benign nodes [54,81] or through fake routing updates and incorrect routing [105]. Once the network is fully segmented with malicious nodes in between the partitions, the attacker can act as a gateway and disrupt the information flow between the network partitions, exclude groups of nodes from the network, or steal peer identities [54,77]. This affects the reliability, autonomy, and connectivity between peers and the CIA properties of P2P networks [72,106,114]. In addition to mounting an eclipse attack by manipulating the overlay network, an attacker that has collected a significant number of peer IDs and acts as a neighbor of benign nodes can easily mount eclipse attacks [54,77,81,107].

Figure 7. Example of an eclipse attack [92]. The attacker successfully segmented the network into 2 ID spaces. The communications between the nodes in the network must be forwarded by the malicious nodes.



Successful eclipse attacks require attackers to possess a high proportion of fake nodes in the network and a higher number of direct routes coming to their nodes than to the average benign nodes in the network [54,77,81], especially in networks with relaxed rules for maintaining the routing table [92]. P2P systems that have no control over node placement in the ID space (eg, Gnutella [54]) or freedom of choice for identifiers (eg, Kad [107]) are highly vulnerable to eclipse attacks. P2P networks are more susceptible to eclipse attacks when they are new [54].

As seen in the Bitcoin network, a botmaster with as few as 24 IP address blocks can eclipse any node with a minimum probability of 85%, irrespective of the number of nodes in the network [114]. Despite new security patches that address eclipse attacks on the Bitcoin network, a novel form of eclipse attack, EREBUS, was found [113], which partitions the network and affects Bitcoin nodes' peering decisions. This shows the likelihood of exploiting eclipses in P2P networks.

Impact of Eclipse Attacks on P2P PHSs

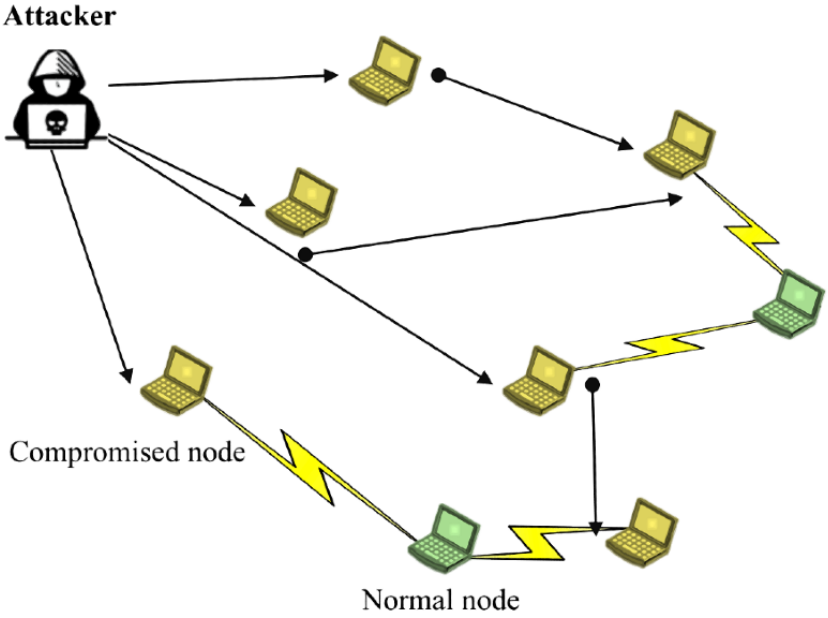
The lack of freedom to select and place identities and the presence of a control infrastructure in centralized and hybrid P2P PHS (eg, our proposed architecture [Figure 3] or the e-toile framework in Switzerland [21]) reduces the impact of any form of eclipse attack on P2P PHSs. This could be higher for decentralized P2P PHSs such as P2P IHE [6,51] because of the absence of centralized trust and control infrastructures and the presence of eclipse attack vectors such as resource routing mechanisms in the network [47]. In addition, a successful attack could allow an attacker to eavesdrop on the conversation between users in the network without potentially compromising the patient's system. P2P PHSs on a patient device can be configured with wearable smart sensors to allow health practitioners or an embedded machine learning model to monitor vital parameters (eg, heart rate variability). In the case of a successful MitM attack on such P2P PHSs, the practitioners or machine learning models may receive unreliable data, which could lead to poor therapeutic or diagnostic decisions and even loss of life [93,135]. An attacker can also share fake messages that an older adult has fallen in order to summon the next-of-kin or emergency services or use the patient's location or personal data for blackmail [93,135].

DDoS Attack

A traditional denial-of-service (DoS) attack stops a service [92,94]. Query flooding is the most common resource and key to mounting DoS on P2P networks [77,105,117]. Invalid or corrupted packets flood the network [95] and impede the delivery of valid requests or messages in the network—byzantine attacks [119]—and therefore stop all communications passing through the affected routes. A DDoS is said to occur when constant streams of invalid packets flood the network in such a way that a single node has to deal with massive traffic and runs out of bandwidth [43,80,81,92]—bandwidth attacks (Figure 8). A lack of central authority can be the root cause for DDoS [97], but the root cause

can also be due to the absence of mechanisms that verify response messages from other nodes (eg, in Kad [98]). Many nodes (or zombies controlled by attackers, where each zombie may control other attacking zombies) participate in DDoS attacks [81,88], while the source of the attack is hidden behind a separate layer or through spoofed IP addresses [84,92,105]. This disguise of the attackers makes it difficult to detect them because they are often only indirectly involved [81].

Figure 8. Example of a distributed denial-of-service (DDoS) attack [92]. The attacker successfully executed the DDoS attack and compromised many nodes in the network. The normal nodes cannot establish connections to other normal nodes.



The previously discussed index and DHT routing table poisoning attacks and file request redirection (or topology change) attacks are other methods of mounting DDoS [77,84,98,102,110,118]. A file request redirection attacker (chatty peer) advertises the possession of many false resources that are rare in the P2P network and then establishes several TCP connections with the victims (requesting peers) [45,100,102]. However, if the requesting peers ask for the blocks of the requesting resource, the attacker only resends handshake messages to the victims and never uploads any blocks. This makes the requesting peers spend much time waiting in vain for the attacker's response and blocking other legitimate users from making connections to them. As such, TCP-connection DDoS comes into effect and affects the availability of entire P2P networks [72]. A request-redirection DDoS attack on internet equipment was used to shut down tech giants' websites (eg, Yahoo and Amazon) in February 2000 [84], which shows the impact severity of DDoS on any network.

DDoS is an active attack that makes it more aggressive. An attacker often attacks the network to prevent certain users from performing their tasks or put the system out of service in one or many segments of the underlying infrastructure [76,84]. The probability of a DDoS attack is high in large P2P networks because nodes have to be reachable (usually outside of firewalls restrictions, etc) by the network [92,117]. Depending on the number of zombies, DDoS on decentralized P2P networks may barely affect the entire network, except for a certain number of affected peers. On the contrary, the impact could be higher on centralized and hybrid systems because communication relies on a single entity that is reachable throughout the network or subnetwork. The higher the number and diversity of nodes involved in the DDoS, the more difficult it is to be blocked [81,97].

Impact of DDoS on P2P PHSs

When P2P PHS providers are hospitals, as in our proposed architecture (Figure 3), and store all patients' medical records, a successful DDoS attack on the network (index or super peers) will have severe consequences. The effect could disrupt the network and data access and cause a delay in treatment and even loss of life (eg, the case of a patient who died after a malware hit a hospital in Germany [130]). In some centralized and hybrid COVID-19 contact tracing systems (eg, PEPP-PT [22] and Trace-Together [23]), the identifiers (ephemeral IDs) that are used to share exposure notifications during smartphone encounters are generated through a central authority (eg, a hospital) and enough of them are generated in batches, for future use and for constructing contact graphs of users when they are infected [136]. A DoS on this server could prevent the IDs and relevant estimations to reach the targets, and the affected persons would have a false sense of safety since they are no longer notified about encountered contacts. In any case, the effect of DDoS is likely higher in centralized and hybrid P2P PHS than in decentralized P2P PHSs such as P2P IHE [6,51]. This is because of the presence of single points that manage other users in the network. However, centralized control mechanisms also ease the tracing of attackers and reduce the probability of DDoS attacks.

P2P Traffic Blockade

In 2008, P2P networks accounted for almost 53% of internet traffic in Germany, followed by web browsing (26%) and streaming (7%) [122]. With respect to P2P network traffic, BitTorrent accounted for 37%, web browsing for 15%, and

eDonkey for 13% of P2P internet traffic [122]. Given the high proportion of P2P traffic in most regions, it is not surprising that a number of internet service providers (ISPs) are using advanced filtering techniques to impose bandwidth limits and throttle or block traffic associated with P2P systems, for instance, by using port numbers, flow features, and deep packet inspections [46,100,121]. In 2012, the United Kingdom High Court ordered, for example, some ISPs (eg, O2, Virgin Media, and TalkTalk) to block BitTorrent P2P traffic owing to its potential for copyright infringements [120].

Impact of P2P Traffic Blockade on P2P PHSs

The consequence of a P2P traffic blockade on any type of P2P PHS could be high because the effect could render the system unavailable over the network, for instance, in a situation where ISPs realize a high proportion of internet traffic caused by P2P networks and impose bandwidth limits or block the traffic. If any P2P PHS user is affected by the blockage, P2P PHSs, for instance, for remote sharing of medical records or COVID-19 exposure notifications will be disrupted. This can potentially affect patient health and contribute to virus spread. As a workaround, users can move to a different region that does not block traffic because P2P systems are not bound by borders. The chances of being affected by a P2P traffic blockade when using a PHS is higher in regions that often use network traffic blockades as a public policy instrument (eg, in authoritarian regimes).

Discussion

Principal Findings

Our findings support the idea that P2P system security is a process rather than a product [33]. Moreover, security encompasses not only technical issues but also human and management problems. Therefore, it is highly relevant for the development and use of P2P PHSs to consider the security issues in P2P networks and the techniques used to exploit them, the security requirements to prevent attacks, peculiarities of attacks, and potential attacker profiles. Our findings are presented in Tables 3 and 4. Security issues such as malware, social engineering attacks, eclipse attacks, DDoS attacks, pollution attacks, and P2P traffic blockades pose high threats (in case of a successful attack) and have a high probability of being exploited in P2P PHSs owing to the high number of factors contributing to their chances of successful exploitation (Table 3); moreover, they can put any P2P PHS out of service, which can potentially affect patients' state of health. For illustrative purposes, we discuss the factors and scores for malware and eclipse attacks in detail below (refer to the section *Identified Security Issues and Their Impact on P2P PHSs* for a detailed discussion of the security issues).

The effect of any malware type depends on its propagation speed and power. Malware that compromised a PHS can be inadvertently spread by the patient (eg, when it is hidden in a patient's health records). Other factors promoting security issues in P2P networks (set-and-forget, no borders, digital winds spreading files, growing use, and network heterogeneity) and no content verification (Table 3) can fuel malware propagation in the network. If attackers compromise super nodes (eg, practitioners or hospital nodes), they can spread malware even more easily. A successful malware attack (eg, Antinny [85,86] or Code-Red [78]) on any P2P PHS can affect the CIA properties of the network and may cause a delay in treatment or even loss of life (eg, the case of a patient who died after a malware hit a hospital in Germany [130]). Malware can attack various network layers (user, network, or transport layers) to mount DoS attacks, poison the network, block P2P traffic, or compromise users' identities or health data.

The severity of malware is low in centralized P2P PHSs (eg, the e-toile framework in Switzerland [21] or P2P PHR [6]; Table 4) because the central index server can simply be used as a trusted computing base [25,26] or a point to deploy control measures to mitigate the propagation of malware in the network. The severity of malware is medium in hybrid P2P PHSs (eg, P2HR [20]), our proposed P2P PHS architecture (Figure 3; Table 4), because there are no central attack profiles, and a federated data ecosystem multiplies the effort required for malware attacks by the number of federated subnetworks. The severity of malware is high in decentralized P2P PHSs (eg, P2P IHE [51]; Table 4) because of the lack of a trusted computing base and high responsibility for individual users to maintain routing information (DHT networks) and security measures [25,26]. Once the neighbor lists of users are infected by malware, the malware can spread further (eg, using a PRS strategy) through the nodes' subnetworks, which contributes to the malware's high propagation speed [78,82].

Factors such as use and network heterogeneity, no borders, and snooping nodes promote the impact of eclipse attacks on P2P networks (Table 3). In most cases, a successful eclipse attack allows an attacker to eavesdrop on the conversation between peers in the network without potentially compromising the victim's system. This impacts the reliability, autonomy, connectivity, and CIA properties of P2P networks [72,106,114].

The severity of eclipse attacks is low in centralized P2P PHS (eg, the e-toile framework in Switzerland [21]; Table 4) because of the difficulty for users to create multiple fake identities (as required to mount an eclipse attack [54,77,81]) and the likely presence of trusted computing infrastructure in centralized P2P PHSs. Nevertheless, attacks on central index servers (or super peers in hybrid P2P PHSs) are likely to be able to snoop network communications. The severity of eclipse attacks is medium in hybrid and decentralized P2P PHSs (eg, P2P IHE [51]) or our proposed P2P PHS architecture [Figure 3; Table 4]), as eclipse attacks require a high number of compromised nodes and are usually achieved through attacks on routing mechanisms [47,54,77,81]. Decentralized and hybrid P2P PHSs require routing mechanisms (eg, DHT) to facilitate health information exchange and communication between patients and practitioners.

Table 3. Factors promoting the security issues.

Security issues	Factors promoting the security issues						
	Inadvertent sensitive information disclosure	Set-and-forget	No borders	Digital winds Spreading Files	Use and network heterogeneity	No content verification	Snooping nodes
Malware	✓[a]	✓	✓	✓	✓	✓	✓
Social engineering attack	✓		✓	✓	✓	✓	✓
Poisoning the network			✓	✓	✓	✓	✓
Sybil attack			✓		✓		✓
Eclipse attack			✓		✓		✓
DDoS attack		✓	✓		✓	✓	✓
Pollution	✓	✓	✓	✓	✓	✓	✓
P2P[b] traffic blockade					✓	✓	✓

[a]Factor present.

[b]P2P: peer-to-peer.

Table 4. Severity ratings for peer-to-peer patient-centered health care information system security.

Security issues	Severity score on P2P PHS[a]			Exemplary security measures
	Centralized	Hybrid	Decentralized	
Malware	Low	Medium	High	• Firewall • Antivirus and antispyware • Mobile agent–based intrusion detection system • Access policies
Social engineering attack	Medium	Medium	Medium	• Education and awareness training
Poisoning the network	Low	Medium	High	• Authentication protocol • Trust and reputation system • Access policies
Sybil attack	Low	Low	Medium	• Authentication protocol • Trust and reputation system • End-to-end encryption
Eclipse attack	Low	Medium	Medium	• Authentication protocol • Trust and reputation system • End-to-end encryption • Access policies
DDoS[b] attack	High	Medium	Medium	• Firewall • Mobile agent–based intrusion detection system • Bandwidth limitation per node • Access policies
Pollution	Low	Medium	Medium	• File and content verification • Trust and reputation system • End-to-end encryption • Removal of polluted content
P2P traffic blockade	High	Medium	Low	• End-to-end encryption • Encryption of P2P traffic

[a]P2P PHS: peer-to-peer patient-centered health care information system.

[b]DDoS: distributed denial-of-service.

Protecting P2P PHSs Against Security Issues

Under normal circumstances, patient-physician relationships are based on trust, and P2P systems generally require trust between their participants [46]. However, uncertainties regarding the protection of user data, single points of failure, and the integrity of the super peers remain. Under our proposed PHS architecture (Figure 3), a trusted registration authority (eg, the German HTI or a hospital) is introduced to the network to handle administrative tasks such as authentication and verification and can also issue or revoke credentials to users based on their behavior [30]. End-to-end encryption [137] can be used to maintain confidentiality in health care information systems [30] and to reduce the trust required for other network participants. For instance, the state-of-the-art cryptographic protocol Signal for end-to-end encryption, which is used by popular instant messaging apps [138], including WhatsApp, Wire, and Facebook Messenger, can be used. Security analyses of the Signal protocol show that it can resist most known attacks [139]. Furthermore, transparency mechanisms can be used to make it easier to hold a provider accountable for violating users' trust [26], for example, certificate transparency can be managed by a set of services and neutral auditors to keep track of X.509 certificates of providers and quickly observe rogue or hacked certificate authorities. Such security techniques reduce the impact of eclipse attacks, DDoS attacks, pollution attacks, poisoning attacks, and P2P traffic blockade on P2P networks [52,81]. For example, an intercepted message can be rendered useless for eclipse attackers by encrypting it.

A discussion of all possible security measures (see Table 4 for examples) for each identified security issue is beyond the scope of this study. In line with the identified security issues, we focus on trust and reputation models (TRM), identity authentication schemes (IAS), and agent-based intrusion detection systems (IDSs). As an overarching guideline, we extended an extant guideline for secure provision of PHSs [2] (Figure 9) with 2 additional steps (*selection and modeling of security measures* [step 3] and *risk assessment* [step 6]). The guideline is useful for supporting individual PHS providers to deal with the complexity of securing P2P PHSs.

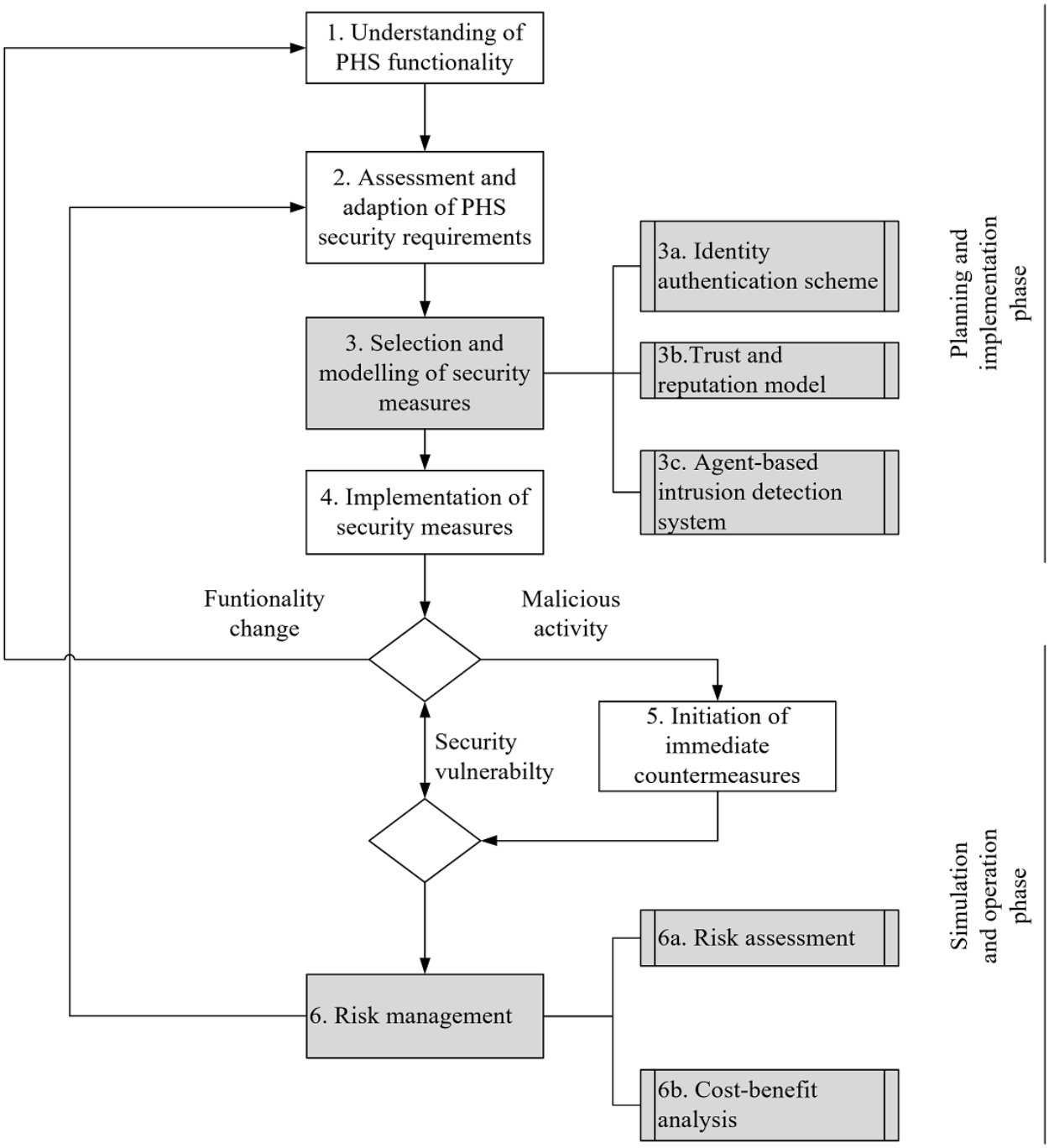
An effective IAS addresses security issues such as Sybil attacks, poisoning attacks, pollution attacks, and MitM attacks [65,81,140] and is essential for health care information systems [2,30]. By authenticating users and resources shared, the origin of pollution or poisoning attacks can be traced, and the attackers can be held accountable. Individual PHS providers leveraging an effective IAS can strengthen security, which has the potential to increase patients' intention to use P2P PHSs. In Germany, the German HTI planned to provide user authentication through smart cards as a security measure for PHS providers [65,141]. However, the introduction of national HTIs often leads to ambiguous, expensive, and protracted projects [65,141]. Until such solutions are widely available, developers of P2P PHSs should consider the use of other IASs for the secure provision of PHSs in public networks [65].

Reputation systems are used to determine the trustworthiness of nodes and to mitigate Sybil, poisoning, pollution, and MitM attacks [142]. Reputation management for resources being shared and peers [143] reduces vulnerabilities such as ID stealth or pseudospoofing [144,145]. TRM techniques can be leveraged in P2P PHS in any situation where a party misbehaves (eg, by supplying inappropriate data to a PHS). Patients can report misbehavior to reputation systems so that it can be reflected in the reputation of the misbehaving party. Polluted resources can also be reported and removed if their reputation is too low [72,73,75,81].

To address the issues of malicious peers, worms, and DDoS attacks in the network, an intelligent mobile agent–based IDS can be deployed in strategic locations (eg, at a hospital node in our proposed P2P PHS architecture, Figure 3; in the DHT network in decentralized P2P PHSs such as P2P IHE [6,51]; or at central index servers of centralized PHSs such as the e-toile PHS [21]) to protect the corresponding subnetworks in P2P networks. There are prototypes of scalable and decentralized agent-based IDS that use 3 types of algorithms (heavy, medium, and light scan algorithms) to detect malicious activities as early as possible [87,146,147]. Backpropagation neural network techniques can be used in IDS to reduce the response times and false alarm rates [148,149]. To improve detection latency and load balancing, a collaborative IDS uses publish and subscribe techniques to selectively route evidence of malicious activities between peers in the network using distributed lookup mechanisms [150,151]. Worms scan and infect certain ports in a network. A firewall can be used to monitor, filter, block, and blacklist them; antivirus and antispyware tools can be leveraged to remove or quarantine any suspicious file [81]. The DDoS can be mitigated by limiting the download bandwidth for each node. Other policies, such as restricting P2P access to verified directories and scanning each file before opening, can mitigate the impact of DDoS, malware, and poisoning attacks [97].

We added risk management (step 6) to the guideline for secure provision of PHSs (Figure 9) to allow for prioritization of security issues with higher impact and for the efficient use of available resources [152]. Risk assessment (step 6a) focuses on the identification and assessment of security issues based on the likelihood of occurrence and the severity of exploits. The cost-benefit analysis involves an analysis of the costs associated with recovering from security breaches. In a situation where the costs for mitigation are higher than the potential impact of a security issue, P2P PHS providers may choose to accept some level of risk.

Figure 9. Guidelines for provision of the patient-centered health care information system (PHS) while ensuring security.



Limitations

This research focuses on security engineering for P2P PHSs. Legal issues with respect to health care security management are beyond the scope of this study. A further limitation of this study is that P2P PHS is an emerging phenomenon; therefore, our study does not provide real-world experiments or a review of past P2P PHS security incidents. Moreover, the bandwidth, computation, and storage cost analyses of the proposed P2P infrastructure, how usability and deployability will affect P2P PHS adoption, and how to handle patient registration with multiple PHS providers are beyond the scope of this study.

Contributions

Our research provides a foundation for understanding P2P system architectures and their advantages and disadvantages. We propose and discuss a federated architecture (Figure 3) suitable for PHS deployment, which could be adopted by any P2P PHS provider, such as insurance companies, hospitals, or other parties who want to implement P2P PHSs while maintaining security. On the basis of the 3 different archetypical P2P system architectures, we elicited and reviewed the inherent security issues and factors promoting the security issues (Table 3). Moreover, we discuss the consequences of the security issues and apply a severity scoring system (Table 4), signifying the impact of each security issue for the 3 different architectures of P2P PHSs—centralized, hybrid, and decentralized—based on

the CVSS definitions (Multimedia Appendix 2). Although a comprehensive discussion of security measures to address each identified security issue is beyond the scope of this study, we offer an overview of potential security measures that are useful for maintaining security in P2P PHSs. We also extended a guideline for the secure provision of PHSs in public networks (Figure 9) for the P2P PHS context [2].

P2P PHSs (eg, COVID-19 contact tracing systems such as PEPP-PT [22] or OnePatient [15]) require research from many perspectives to facilitate widespread use because they are an emerging phenomenon, pose major security issues (eg, by requiring patients to manage information security largely by themselves [65]), and are understudied. Extant research on PHS security, privacy, and end-user features [2,28-31] focuses on centralized and DLT-based PHS. Our research serves as an introduction to P2P PHSs and potential security issues and countermeasures. From an ethical perspective, our study is of interest to initiatives aimed at empowering patients to take ownership of and control access to their health data. P2P PHSs promote socially desirable design features such as openness, reduced dependence on platforms, abandonment of data silos, and secure patient-to-practitioner communication. Given that the security challenges are appropriately addressed, P2P PHSs are also promising for simplifying the implementation of data protection principles (eg, GDPR [8,34]). Secure P2P PHSs will not only attract more stakeholders but will also be more efficient in achieving the goals of patient-centered digital ecosystems [153].

Future Research

Opportunities for future research include improved designs of security models, such as IAS, TRM, and intelligent mobile agent–based IDS, to strengthen security. PHSs have other more safety-related security requirements that should also be incorporated in their design, such as emergency access and guardian support. Such features are vital for P2P PHS to facilitate access in situations where patients are incapacitated. However, they are also likely to invoke privacy concerns and data protection challenges, as they require access to sensitive information without consulting patients. By using reliable and patient-centered backup options, P2P PHS providers can integrate identity authentication management in backup servers to facilitate the replacement of patient credentials in a situation where they lose access to their credentials (eg, a stolen laptop). In addition to the development of approaches to improve education and awareness of patients regarding information security challenges inherent to the sharing of data with third parties [8], a questionnaire-based study focusing on other P2P PHS stakeholders and asking about their security and privacy concerns with respect to P2P PHSs could yield valuable contributions. A guideline for the evaluation of P2P PHSs based on information security standards (eg, ISO 27799:2016) could also be very useful.

Conclusions

The idea of P2P PHSs to break up barriers among patients, health care systems, physicians, and other stakeholders is appealing. From the patients' perspective, being empowered to conveniently take ownership of and control access to their health data through PHS might bring forth a digital ecosystem that makes patients a more active contributor in their own care and can streamline health care activities such as receiving and accurately interpreting laboratory test results. In the United States, HIPAA [6] specifies that patients have the liberty "to see and get copies of their records, and request amendments"; however, the act does not go into detail on appropriate approaches to give access [3,30,154]. Currently, PHSs use DLT, P2P technology, or centralized databases for deployment. To mitigate the impact of security issues in centralized databases and the lack of fit of DLT with PHS use cases, P2P PHSs emerged (eg, OnePatient [15], doc.ai brands [7], or COVID-19 proximity tracing systems such as Stoop [24]), which store health records locally (on any patient edge device such as a mobile phone, a tablet computer, a desktop computer, etc) under the control of individual device owners.

The benefits of P2P networks for PHSs include more options for privacy self-management, autonomous control of infrastructure, and high availability. However, these advantages are associated with complications, as patients must also manage information security largely by themselves. Gartner claims that costs for remediating security issues would be reduced by 75% if only 50% of system vulnerabilities were detected and remediated before production [155]. Building a successful P2P system that does not result in privacy or security violations for users is difficult [26] and entails a collective effort that fixes the remaining problems (eg, absence of a centralized entity to detect malicious attacks and increased chances of exposing network traffic patterns) with clear considerations of network security and ease of use.

The enormous value of health data requires the provision of security measures to protect PHSs from cyberattacks. Overcoming security and privacy barriers in P2P PHS is also important for increasing patients' intention to use PHSs. PHS providers and developers should neither ignore the inherent or past security issues of P2P systems nor be careless about future ones.

Acknowledgments

This work was supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF) and by KASTEL Security Research Labs.

Conflicts of Interest

JG is a chief executive officer at Refinio GmbH, a company that provides peer-to-peer patient-centered health care information systems (eg, OnePatient). IAY was involved in weekly discussions with software developers working on OnePatient patient-centered

health care information system about peer-to-peer patient-centered health care information systems innovations, motivations, contributions, and foundational architectures for this research. BE reports grants from adidas AG, outside the submitted work. In addition, BE has a patent related to gait assessment pending and reports ownership of Portabiles GmbH and Portabiles HealthCare Technologies GmbH.

Multimedia Appendix 1
List of individual journals and conferences.
[DOCX File , 27 KB - jmir_v23i11e24460_app1.docx]

Multimedia Appendix 2
Definition of Consequence of Exploitation. The rate estimation was guided by the Common Vulnerability Scoring System which provides a way to capture the principal characteristics of a vulnerability and produce a numerical score reflecting its severity.
[DOCX File , 22 KB - jmir_v23i11e24460_app2.docx]

References

1. Healthcare of the Future: the digital revolution of the healthcare sector - ecosystem, use cases, benefits, challenges and recommendations for action. Porsche-Consulting. 2018. URL: https://www.porsche-consulting.com/fileadmin/docs/Startseite/News/tt1162/Porsche_Consulting_Studie_Healthcare_of_the_Future_EN.pdf [accessed 2019-09-12]
2. Dehling T, Sunyaev A. Secure provision of patient-centered health information technology services in public networks—leveraging security and privacy features provided by the German nationwide health information technology infrastructure. Electron Markets 2014 Feb 8;24(2):89-99 [FREE Full text] [doi: 10.1007/s12525-013-0150-6]
3. Sunyaev A, Chornyi D, Mauro C, Krcmar H. Evaluation framework for personal health records: Microsoft Healthvault vs. Google Health. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. 2010 Presented at: 43rd Hawaii International Conference on System Sciences; Jan. 5-8, 2010; Honolulu, HI, USA p. 1-10. [doi: 10.1109/hicss.2010.192]
4. Krist AH, Woolf SH. A vision for patient-centered health information systems. J Am Med Assoc 2011 Jan 19;305(3):300-301 [FREE Full text] [doi: 10.1001/jama.2010.2011] [Medline: 21245186]
5. Dehling T, Sunyaev A. Architecture and design of a patient-friendly eHealth web application: patient information leaflets and supplementary services. In: Proceedings of the 18th Americas Conference on Information Systems (AMCIS 2012). 2012 Presented at: 18th Americas Conference on Information Systems (AMCIS 2012); August 9-11, 2012; Seattle, Washington URL: https://ssrn.com/abstract=2152925
6. Horne WC, Miled ZB. Making the case for a P2P personal health record. Information 2020 Oct 31;11(11):512. [doi: 10.3390/info11110512]
7. Doc.AI brands: Passport, Serenity, Doc.ai, and Genewall. Doc.ai. 2020. URL: https://doc.ai/ [accessed 2019-10-01]
8. Hager A, Lindblad S, Brommels M, Salomonsson S, Wannheden C. Sharing patient-controlled real-world data through the application of the theory of commons: action research case study. J Med Internet Res 2021 Jan 19;23(1):e16842 [FREE Full text] [doi: 10.2196/16842] [Medline: 33464212]
9. Sinthanayothin C, Bholsithi W, Wongwaen N, Xuto P. ZBaby: Android application for pregnancy due date, fetus development simulation and weight gain during pregnancy. In: Proceedings of the International Computer Science and Engineering Conference (ICSEC). 2014 Presented at: International Computer Science and Engineering Conference (ICSEC); July 30 - Aug. 1, 2014; Khon Kaen, Thailand. [doi: 10.1109/icsec.2014.6978130]
10. Spitzer J. 63% of Americans don't know where their medical data is stored: 8 survey insights. Becker's Healthcare. 2018. URL: https://www.beckershospitalreview.com/ehrs/63-of-americans-don-t-know-where-their-medical-data-is-stored-8-survey-insights.html [accessed 2018-10-21]
11. MedicalChain Whitepaper version 2.1. MedicalChain. 2018. URL: https://medicalchain.com/Medicalchain-Whitepaper-EN.pdf [accessed 2018-12-05]
12. HealthBank. 2019. URL: https://www.healthbank.coop/ [accessed 2019-01-06]
13. Zhou L, DeAlmeida D, Parmanto B. Applying a user-centered approach to building a mobile personal health record app: development and usability study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]
14. Aligning stakeholders in a new healthcare ecosystem. MintHealth. 2018. URL: https://www.minthealth.io/wp-content/uploads/2018/04/MH_WHITEPAPER_04.18.pdf [accessed 2019-03-05]
15. OnePatient. RefinioONE. 2019. URL: https://refinio.net/software.html [accessed 2018-10-10]
16. Kleinke JD. Dot-gov: market failure and the creation of a national health information technology system. Health Aff (Millwood) 2005;24(5):1246-1262. [doi: 10.1377/hlthaff.24.5.1246] [Medline: 16162569]
17. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data (OBD). 2016 Presented at: 2nd International Conference on Open and Big Data (OBD); Aug. 22-24, 2016; Vienna, Austria p. 25-30. [doi: 10.1109/obd.2016.11]

18. Beinke JH, Fitte C, Teuteberg F. Towards a stakeholder-oriented blockchain-based architecture for electronic health records: design science research study. J Med Internet Res 2019 Oct 07;21(10):e13585 [FREE Full text] [doi: 10.2196/13585] [Medline: 31593548]
19. Wright CS. Bitcoin: a peer-to-peer electronic cash system. SSRN J 2020:9986. [doi: 10.2139/ssrn.3440802]
20. King Z. P2HR, a personalized condition-driven person health record. Master thesis submitted to the Faculty of Purdue University, Indianapolis, Indiana. 2017. URL: https://scholarworks.iupui.edu/handle/1805/13604 [accessed 2019-10-01]
21. Geissbuhler A, Spahni S, Assimacopoulos A, Raetzo M, Gobet G. Design of a patient-centered, multi-institutional healthcare information network using peer-to-peer communication in a highly distributed architecture. Stud Health Technol Inform 2004;107(Pt 2):1048-1052. [Medline: 15360972]
22. Troncoso C, Payer M, Hubaux J, Salathé M, Larus J, Bugnion E, et al. Decentralized privacy-preserving proximity tracing - GitHub DP-3T Documents. GitHub. 2020. URL: https://github.com/DP-3T/documents [accessed 2020-07-01]
23. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: privacy considerations and related trade-offs. arXiv. 2020. URL: https://arxiv.org/abs/2003.11511 [accessed 2020-04-01]
24. Ulrich B, Bernauer A, Blocher M, Gollatz B, Judmayer A, Koppmann M, et al. Technical and legal review of the stopp corona app by the Austrian Red Cross. NOYB – European Center for Digital Rights. 2020. URL: https://noyb.eu/sites/default/files/2020-04/report_stopp_corona_app_english_v1.0_0.pdf [accessed 2020-05-01]
25. Rushby J. Design and verification of secure systems. SIGOPS Oper Syst Rev 1981 Dec;15(5):12-21. [doi: 10.1145/1067627.806586]
26. Troncoso C, Isaakidis M, Danezis G, Halpin H. Systematizing decentralization and privacy: lessons from 15 years of research and deployments. Proc Privacy Enhanc Technol 2017;4:404-426 [FREE Full text] [doi: 10.1515/popets-2017-0056]
27. Solid project by Tim Berners-Lee. Solid Project. 2020. URL: https://solidproject.org/ [accessed 2020-02-01]
28. Kaletsch A, Sunyaev A. Privacy engineering: personal health records in cloud computing environments. In: Proceedings of the International Conference on Information Systems, ICIS 2011. 2011 Presented at: International Conference on Information Systems, ICIS 2011; December 4-7, 2011; Shanghai, China, URL: https://www.researchgate.net/publication/221600139_Privacy_Engineering_Personal_Health_Records_in_Cloud_Computing_Environments
29. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
30. Fernández-Alemán JL, Señor IC, Lozoya PA, Toval A. Security and privacy in electronic health records: a systematic literature review. J Biomed Inform 2013 Jun;46(3):541-562 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.003] [Medline: 23305810]
31. Müthing J, Brüngel R, Friedrich CM. Server-focused security assessment of mobile health apps for popular mobile platforms. J Med Internet Res 2019 Jan 23;21(1):e9818 [FREE Full text] [doi: 10.2196/jmir.9818] [Medline: 30672738]
32. Farn K, Hwang J, Lin S. Study on applying ISO/DIS 27799 to healthcare industry's ISMS. WSEAS Transactions on Biology and Biomedicine. 2007. URL: https://wseas.org/wseas/cms.action?id=4011 [accessed 2020-09-01]
33. Schneier B. In: Phil S, editor. Applied Cryptography: Protocols, Algorithms, Source Code in C. New York, United States: John Wiley & Sons; 2007.
34. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mhealth applications: a scoping review. IEEE Access 2020;8:104247-104268. [doi: 10.1109/access.2020.2999934]
35. Vu QH, Lupu M, Ooi BC. Architecture of peer-to-peer systems. In: Peer-to-Peer Computing. Berlin: Springer; 2010:11-37.
36. Vakili G, Khorsandi S. Engineering a peer to peer architecture: a complex adaptive system approach. In: Proceedings of the IEEE International Systems Conference. 2010 Presented at: IEEE International Systems Conference; April 5-8, 2010; San Diego, CA, USA p. 520-523. [doi: 10.1109/systems.2010.5482487]
37. Records, computers, and the rights of citizens: Report of the US Dept Welfare Secretary's Advisory Committee on Automated Personal Data Systems. US Department of Health, Education and Welfare. 1973. URL: https://www.justice.gov/opcl/docs/rec-com-rights.pdf [accessed 2020-05-05]
38. Gassmann H. OECD guidelines governing the protection of privacy and transborder flows of personal data. Comput Networks (1976) 1981 Apr;5(2):127-141 [FREE Full text] [doi: 10.1016/0376-5075(81)90068-4]
39. Notifiable Data Breaches Report: July–December 2019. The Office of the Australian Information Commissioner (OAIC). 2020. URL: https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-statistics/notifiable-data-breaches-report-july-december-2019/ [accessed 2020-07-01]
40. Protenus 2019 Breach Barometer: 15m+ patient records breached in 2018 as hacking incidents continue to climb. Protenus, Inc. in Collaboration with DataBreaches.net. 2019. URL: https://email.protenus.com/hubfs/Breach_Barometer/2018/2019%20Breach%20Barometer%20Annual%20Report.pdf [accessed 2020-04-01]
41. Burton C, De Boel L, Kuner C. The final European Union general data protection regulation. Wilson Sonsini. 2016. URL: https://www.wsgr.com/en/insights/the-final-european-union-general-data-protection-regulation.html [accessed 2020-06-02]
42. Kelly G, McKenzie B. Security, privacy, and confidentiality issues on the internet. J Med Internet Res 2002;4(2):E12. [doi: 10.2196/jmir.4.2.e12] [Medline: 12554559]

43. Naoumov N, Ross K. Exploiting P2P systems for DDoS attacks. In: Proceedings of the 1st international conference on Scalable information systems. 2006 Presented at: InfoScale '06: 1st International Conference on Scalable Information Systems; May 30-June 1, 2006; Hong Kong, China p. 47. [doi: 10.1145/1146847.1146894]
44. Jan MA, Nanda P, He X, Liu RP. A Sybil attack detection scheme for a forest wildfire monitoring application. Future Generation Comput. Syst 2018 Mar;80:613-626 [FREE Full text] [doi: 10.1016/j.future.2016.05.034]
45. Nwebonyi FN, Martins R, Correia ME. Reputation based approach for improved fairness and robustness in P2P protocols. Peer-to-Peer Netw Appl 2018 Dec 6;12(4):951-968 [FREE Full text] [doi: 10.1007/s12083-018-0701-x]
46. Wallach DS. A survey of peer-to-peer security issues. In: Proceedings of the International Symposium on Software Security. 2003 Presented at: International Symposium on Software Security; November 4-6, 2003; Tokyo, Japan p. 42-57. [doi: 10.1007/3-540-36532-x_4]
47. Ismail H. Analyzing and mitigating security threats in P2P systems. Darmstadt, Technische Universität, [Ph.D. Thesis]. 2018. URL: https://tuprints.ulb.tu-darmstadt.de/7812/ [accessed 2019-10-04]
48. Steve C. Host software - RFC-1. Datacracker - IETF. 1969. URL: https://tools.ietf.org/html/rfc1 [accessed 2019-06-02]
49. Doyle JF. Peer-to-peer: harnessing the power of disruptive technologies. Ubiquity 2001 May;2001(May):2. [doi: 10.1145/375348.377427]
50. Shen X, Yu H, Buford J, Akon M. Handbook of Peer-to-Peer Networking. Boston, MA: Springer; 2010:1-1500.
51. Urovi V, Olivieri AC, Bromuri S, Fornara N, Schumacher MI. A peer to peer agent coordination framework for IHE based cross-community health record exchange. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing. 2013 Presented at: SAC '13: 28th Annual ACM Symposium on Applied Computing; March 18 - 22, 2013; Coimbra Portugal p. 1355-1362. [doi: 10.1145/2480362.2480617]
52. Palomar E, Estevez-Tapiado J, Hernandez-Castro J, Ribagorda A. Security in P2P networks: survey and research directions. In: Emerging Directions in Embedded and Ubiquitous Computing. Berlin: Springer; 2006:183-192.
53. Kurose JF. Computer Networking: A Top-Down Approach (6th Edition). London, U.K: Pearson; 2012:1-864.
54. López-Fuentes FD, Eugui-De-Alba I, Ortíz-Ruiz OM. Evaluating P2P networks against eclipse attacks. Procedia Technol 2012;3:61-68. [doi: 10.1016/j.protcy.2012.03.007]
55. Balakrishnan H, Kaashoek MF, Karger D, Morris R, Stoica I. Looking up data in P2P systems. Commun ACM 2003 Feb;46(2):43-48. [doi: 10.1145/606272.606299]
56. De S, Barik MS, Banerjee I. Goal based threat modeling for peer-to-peer cloud. Procedia Comput Sci 2016;89:64-72. [doi: 10.1016/j.procs.2016.06.010]
57. Feldotto M, Scheideler C, Graffi K. A self-stabilizing overlay network for nodes with heterogeneous bandwidths. In: Proceedings of the 14th IEEE International Conference on Peer-to-Peer Computing. 2014 Presented at: 14th IEEE International Conference on Peer-to-Peer Computing; Sept. 8-12, 2014; London, UK p. 1-10. [doi: 10.1109/p2p.2014.6934300]
58. Sanchez-Artigas M, Garcia-Lopez P. On routing in distributed hash tables: is reputation a shelter from malicious behavior and churn? In: Proceedings of the IEEE Ninth International Conference on Peer-to-Peer Computing. 2009 Presented at: IEEE Ninth International Conference on Peer-to-Peer Computing; Sept. 9-11, 2009; Seattle, WA, USA p. 31-40. [doi: 10.1109/p2p.2009.5284546]
59. Wang Q, Mittal P, Borisov N. In search of an anonymous and secure lookup: attacks on structured peer-to-peer anonymous communication systems. In: Proceedings of the 17th ACM Conference on Computer and Communications Security. 2010 Presented at: CCS '10: 17th ACM Conference on Computer and Communications Security 2010; Oct 4-8, 2010; Chicago Illinois USA p. 308-318. [doi: 10.1145/1866307.1866343]
60. Omicini A, Denti E. From tuple spaces to tuple centres. Sci Comput Program 2001 Nov;41(3):277-294. [doi: 10.1016/s0167-6423(01)00011-9]
61. Chen G, Hu T, Jiang D, Lu P, Tan K, Vo HT, et al. Bestpeer++: a peer-to-peer based large-scale data processing platform. IEEE Trans Knowl Data Eng 2014 Jun;26(6):1316-1331. [doi: 10.1109/tkde.2012.236]
62. Decker C, Eidenbenz R, Wattenhofer R. Exploring and improving BitTorrent topologies. In: Proceedings of the IEEE P2P 2013 Conference. 2013 Presented at: IEEE P2P 2013 Conference; Sept. 9-11, 2013; Trento, Italy. [doi: 10.1109/P2P.2013.6688698]
63. de Leon DC, Stalick AQ, Jillepalli AA, Haney MA, Sheldon FT. Blockchain: properties and misconceptions. Asia Pcfc Jrnl Innvtn & Entrprnshp 2017 Dec 04;11(3):286-300. [doi: 10.1108/apjie-12-2017-034]
64. Donnelly N. The PACE system: a P2P architecture for cloud based EHealth systems 2015. Master of Science thesis, Dublin City University. 2015. URL: http://doras.dcu.ie/20781/ [accessed 2020-02-04]
65. Abdullahi I, Dehling T, Kluge F, Eskofier B, Sunyaev A. Online at will: a novel protocol for mutual authentication in peer-to-peer networks for patient-centered health care information systems. In: Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021). 2021 Presented at: 54th Hawaii International Conference on System Sciences (HICSS 2021); Jan 05, 2021; Kauai, Hawaii, USA p. 3828. [doi: 10.24251/hicss.2021.463]
66. Yang B, Garcia-Molina H. Improving search in peer-to-peer networks. In: Proceedings 22nd International Conference on Distributed Computing Systems. 2002 Jul Presented at: 22nd International Conference on Distributed Computing Systems; July 2-5, 2002; Vienna, Austria p. 5-14 URL: https://ieeexplore.ieee.org/document/1022237 [doi: 10.1109/ICDCS.2002.1022237]

67. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
68. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report, Keele University and University of Durham. 2007. URL: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf [accessed 2019-10-04]
69. Friese S. Qualitative Data Analysis with ATLAS. Thousand Oaks, CA: SAGE Publications Ltd; 2019:1-344.
70. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
71. Lu M, Lee P, Lui J. Identity attack and anonymity protection for P2P-VoD systems. In: Proceedings of the Nineteenth IEEE International Workshop on Quality of Service. 2011 Presented at: Nineteenth IEEE International Workshop on Quality of Service; June 6-7, 2011; San Jose, CA, USA. [doi: 10.1109/iwqos.2011.5931313]
72. Gheorghe G, Lo Cigno R, Montresor A. Security and privacy issues in P2P streaming systems: a survey. Peer-to-Peer Netw Appl 2010 Apr 23;4(2):75-91 [FREE Full text] [doi: 10.1007/s12083-010-0070-6]
73. Wang JH, Wang C, Yang J, An C. A study on key strategies in P2P file sharing systems and ISPs' P2P traffic management. Peer-to-Peer Netw Appl 2011 Jan 7;4(4):410-419 [FREE Full text] [doi: 10.1007/s12083-010-0098-7]
74. Zhang P, Helvik BE. Towards green P2P: understanding the energy consumption in P2P under content pollution. In: Proceedings of the IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing. 2010 Presented at: IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing; Dec. 18-20, 2010; Hangzhou, China. [doi: 10.1109/greencom-cpscom.2010.45]
75. de Almeida RB, Natif JA, da Silva AP, Vieira AM. Pollution and whitewashing attacks in a P2P live streaming system: analysis and counter-attack. In: Proceedings of the IEEE International Conference on Communications (ICC). 2013 Presented at: IEEE International Conference on Communications (ICC); June 9-13, 2013; Budapest, Hungary. [doi: 10.1109/icc.2013.6654819]
76. Tang H, Yang Z, Chen L. Analysis on security issues for mobile P2P networks. In: Proceedings of the 8th International Conference on Wireless Communications, Networking and Mobile Computing. 2012 Presented at: 8th International Conference on Wireless Communications, Networking and Mobile Computing; Sept. 21-23, 2012; Shanghai, China. [doi: 10.1109/WICOM.2012.6478429]
77. Yue X, Qiu X, Ji Y, Zhang C. P2P attack taxonomy and relationship analysis. In: Proceedings of the 11th International Conference on Advanced Communication Technology. 2009 Presented at: 11th International Conference on Advanced Communication Technology; Feb. 15-18, 2009; Gangwon, Korea (South) URL: https://ieeexplore.ieee.org/document/4809630?arnumber=4809630
78. Yu W, Chellappan S, Wang X, Xuan D. Peer-to-peer system-based active worm attacks: modeling, analysis and defense. Comput Commun 2008 Nov;31(17):4005-4017. [doi: 10.1016/j.comcom.2008.08.008]
79. Seedorf J. Security issues for P2P-based voice-and video-streaming applications. In: Camenisch J, Kesdogan D, editors. iNetSec 2009 - Open Research Problems in Network Security. IFIP Advances in Information and Communication Technology. Berlin: Springer; 2009:95-110.
80. Washbourne L. A survey of P2P network security. arXiv. 2015. URL: https://arxiv.org/abs/1504.01358 [accessed 2019-10-05]
81. Yang Y, Yang L. A survey of peer-to-peer attacks and counter attacks. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 2012. URL: http://worldcomp-proceedings.com/proc/p2012/SAM9754.pdf [accessed 2018-10-04]
82. Chunyan X, Zhiyu Y. The research of worms in P2P networks. In: Proceedings of the International Conference on Computational Intelligence and Natural Computing. 2009 Presented at: International Conference on Computational Intelligence and Natural Computing; June 6-7, 2009; Wuhan, China. [doi: 10.1109/cinc.2009.248]
83. Sokolova M, El Emam K, Arbuckle L, Neri E, Rose S, Jonker E. P2P watch: personal health information detection in peer-to-peer file-sharing networks. J Med Internet Res 2012 Jul 09;14(4):e95 [FREE Full text] [doi: 10.2196/jmir.1898] [Medline: 22776692]
84. Abdelouahab M, Bouabdallah A, Achemlal M, Laniepce S. The Topology change attack: threat and impact. J Universal Comput Sci 2009;15(2):465-487 [FREE Full text]
85. Johnson ME, McGuire D, Willey ND. The evolution of the peer-to-peer file sharing industrythe security risks for users. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). 2008 Presented at: 41st Annual Hawaii International Conference on System Sciences (HICSS 2008); Jan. 7-10, 2008; Waikoloa, HI, USA p. 7-10. [doi: 10.1109/hicss.2008.436]
86. Kang M. Responsive Security: Be Ready to be Secure. Boca Raton: CRC Press; 2014:1-259.
87. Zhang Y, Li Z, Hu Z, Tu H, Lin H. A P2P e-commerce related network security issue: P2P worm. In: Proceedings of the International Symposium on Electronic Commerce and Security. 2008 Presented at: International Symposium on Electronic Commerce and Security; Aug. 3-5, 2008; Guangzhou, China. [doi: 10.1109/isecs.2008.187]
88. Barse Y, Tidke S. A study on BOTNET attacks and detection techniques. IOSR J Electri Electron Eng (IOSR-JEEE) 2020;15(3):1-5 [FREE Full text] [doi: 10.9790/1676-1503020105]

89. Wang P, Wu L, Aslam B, Zou CC. Analysis of peer-to-peer botnet attacks and defenses. In: Król D, Fay D, Gabryś B, editors. Propagation Phenomena in Real World Networks. Intelligent Systems Reference Library. Cham: Springer; 2015.
90. Jang-Jaccard J, Nepal S. A survey of emerging threats in cybersecurity. J Comput Syst Sci 2014 Aug;80(5):973-993 [FREE Full text] [doi: 10.1016/j.jcss.2014.02.005]
91. Milpitas C. SonicWALL sees dramatic jump in IoT malware, encrypted threats, web app attacks through third quarter. SonicWALL. 2019. URL: https://www.sonicwall.com/news/dramatic-jump-in-iot-malware-encrypted-threats-web-app-attacks-third-quarter/ [accessed 2020-06-04]
92. Engle M, Khan J. Vulnerabilities of P2P systems and a critical look at their solutions. Technical Report, Kent State University. 2006. URL: https://www.semanticscholar.org/paper/Vulnerabilities-of-P2P-Systems-and-a-Critical-Look-Engle-Khan/4f86c19fd5b8147cbc5da222903a8c48e79cf7f8 [accessed 2018-10-01]
93. Ali MS, Vecchio M, Putra GD, Kanhere SS, Antonelli F. A decentralized peer-to-peer remote health monitoring system. Sensors (Basel) 2020 Mar 16;20(6):s20061656 [FREE Full text] [doi: 10.3390/s20061656] [Medline: 32188135]
94. Schäfer J, Malinka K. Security in peer-to-peer networks: empiric model of file diffusion in BitTorrent. In: Proceedngs of the Fourth International Conference on Internet Monitoring and Protection. 2009 Presented at: Fourth International Conference on Internet Monitoring and Protection; May 24-28, 2009; Venice/Mestre, Italy p. 39-44. [doi: 10.1109/icimp.2009.14]
95. Kamat P, Gite S, Kumar M, Patil S. A critical analysis of P2P communication, security concerns and solutions. Int J Appl Eng Res 2014;9:30899-30909 [FREE Full text]
96. Lu C, Li F, Cheng J, Ni B, Li H. Quantitative analysis and countermeasures research of P2P routing attacks. In: Proceedings of the 2nd International Workshop on Intelligent Systems and Applications. 2010 Presented at: 2nd International Workshop on Intelligent Systems and Applications; May 22-23, 2010; Wuhan, China. [doi: 10.1109/iwisa.2010.5473277]
97. Qi M. P2P network-targeted DDoS attacks. In: Proceedings of the Second International Conference on the Applications of Digital Information and Web Technologies. 2009 Presented at: Second International Conference on the Applications of Digital Information and Web Technologies; Aug. 4-6, 2009; London, UK. [doi: 10.1109/icadiwt.2009.5273837]
98. Koo H, Lee Y, Kim K, Roh B, Lee C. A DDoS attack by flooding normal control messages in Kad P2P networks. In: Proceedings of the 14th International Conference on Advanced Communication Technology (ICACT). 2012 Presented at: 14th International Conference on Advanced Communication Technology (ICACT); Feb. 19-22, 2012; PyeongChang, Korea (South) p. 19-22 URL: https://ieeexplore.ieee.org/document/6174645
99. Cuevas R, Kryczka M, Cuevas A, Kaune S, Guerrero C, Rejaie R. Is content publishing in BitTorrent altruistic or profit-driven? In: Proceedings of the 6th International COnference. 2010 Presented at: Co-NEXT '10: Conference on emerging Networking EXperiments and Technologies; Nov. 30 - Dec. 03, 2010; Philadelphia, Pennsylvania p. 1-12. [doi: 10.1145/1921168.1921183]
100. Divac-Krnic L, Ackermann R. 31 security-related issues in peer-to-peer networks. In: Peer-to-Peer Systems and Applications. Berlin: Springer; 2005:529-545.
101. Yuan Q, Little A, Kabore M, Kabore Y. A study of index poisoning in peer-to-peer file sharing systems. Int J Cybernet Informatics 2014 Dec 31;3(6):11-24. [doi: 10.5121/ijci.2014.3602]
102. Lee Y, Kim K, Roh BH. DDoS attack by file request redirection in Kad P2P network. In: Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. 2012 Presented at: International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery; Oct. 10-12, 2012; Sanya, China. [doi: 10.1109/cyberc.2012.91]
103. Karame GO, Androulaki E, Capkun S. Double-spending fast payments in bitcoin. In: Proceedings of the 2012 ACM conference on Computer and Communications Security. 2012 Presented at: CCS'12: the ACM Conference on Computer and Communications Security; October 16 - 18, 2012; Raleigh North Carolina USA p. 906-917. [doi: 10.1145/2382196.2382292]
104. McConaghy T, Marques R, Müller A. BigchainDB: a scalable blockchain database - Whitepaper. BigChainDB. 2016. URL: https://www.bigchaindb.com/whitepaper/ [accessed 2019-10-04]
105. Touceda DS, Sierra JM, Izquierdo A, Schulzrinne H. Survey of attacks and defenses on P2PSIP communications. IEEE Commun Surv Tutorials 2011:00152. [doi: 10.1109/surv.2011.060711.00152]
106. Tselios C, Birkos K, Galiotos P, Kotsopoulos S, Dagiuklas T. Malicious threats and novel security extensions in P2PSIP. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops. 2012 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops; 19-23 March, 2012; Lugano, Switzerland p. 19-23. [doi: 10.1109/percomw.2012.6197612]
107. Kohnen M, Leske M, Rathgeb EP. Conducting and optimizing eclipse attacks in the Kad peer-to-peer network. In: Proceedings of the International Conference on Research in Networking. 2009 Presented at: International Conference on Research in Networking; May 11-15, 2009; Aachen, Germany p. 104-116. [doi: 10.1007/978-3-642-01399-7_9]
108. Zou W, Zhang Y, Zhang J, Zhou M, Liu B. Survey of eclipse attacks on DHT networks. Qinghua Daxue Xuebao / J Tsinghua Univ 2011;51(10):1306-1311 [FREE Full text]
109. Aiello LM, Milanesio M, Ruffo G, Schifanella R. An identity-based approach to secure P2P applications with Likir. Peer-to-Peer Netw Appl 2011 Jan 6(4):420-438 [FREE Full text] [doi: 10.1007/s12083-010-0099-6]

110. Mi W, Qiu X, Zhang C. The analysis of security threats in structured P2P load balancing schemes. In: Proceedings of the International Conference on Cloud and Service Computing. 2011 Presented at: International Conference on Cloud and Service Computing; Dec. 12-14, 2011; Hong Kong, China. [doi: 10.1109/csc.2011.6138537]
111. Douceur J. The sybil attack. In: Proceedings of the International Workshop on Peer-to-Peer Systems. 2002 Presented at: International Workshop on Peer-to-Peer Systems; March 7-8, 2002; Cambridge, MA, USA p. 251-260. [doi: 10.1007/3-540-45748-8_24]
112. Kannengießer N, Lins S, Dehling T, Sunyaev A. Trade-offs between distributed ledger technology characteristics. ACM Comput Surv 2020 Jul;53(2):1-37. [doi: 10.1145/3379463]
113. Tran M, Choi I, Moon GJ, Vu AV, Kang MS. A stealthier partitioning attack against bitcoin peer-to-peer network. In: Proceedings of the IEEE Symposium on Security and Privacy (SP). 2020 Presented at: IEEE Symposium on Security and Privacy (SP); May 18-21, 2020; San Francisco, CA, USA. [doi: 10.1109/sp40000.2020.00027]
114. Heilman E, Kendler A, Zohar A. Eclipse attacks on bitcoin's peer-to-peer network. In: Proceedings of the 24th USENIX Security Symposium (USENIX Security 15). 2015 Presented at: 24th USENIX Security Symposium (USENIX Security 15); August 12-14, 2015; Washington, D. C URL: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/heilman
115. Vanhoef M, Piessens F. Key reinstallation attacks: forcing nonce reuse in WPA2. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017 Presented at: CCS '17: 2017 ACM SIGSAC Conference on Computer and Communications Security; October 30 - November 3, 2017; Dallas Texas USA p. 1313-1328. [doi: 10.1145/3133956.3134027]
116. Yoon S, Park S, Park H, Yoo HS. Security analysis of vulnerable Wi-Fi Direct. In: Proceedings of the 8th International Conference on Computing and Networking Technology (INC, ICCIS and ICMIC). 2012 Presented at: 8th International Conference on Computing and Networking Technology (INC, ICCIS and ICMIC); Aug. 27-29, 2012; Gyeongju, Korea (South) URL: https://ieeexplore.ieee.org/document/6418681
117. Dejmal S, Fern A, Nguyen T. Reinforcement learning for vulnerability assessment in peer-to-peer networks. In: Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008. 2008 Presented at: Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008; July 13-17, 2008; Chicago, Illinois, USA p. 1655-1662 URL: https://www.researchgate.net/publication/220269202_Reinforcement_Learning_for_Vulnerability_Assessment_in_Peer-to-Peer_Networks
118. Tong J, Xiong G. A research on the vulnerability in popular P2P protocols. In: Proceedings of the 8th International Conference on Communications and Networking in China (CHINACOM). 2013 Presented at: 8th International Conference on Communications and Networking in China (CHINACOM); Aug. 14-16, 2013; Guilin, China. [doi: 10.1109/chinacom.2013.6694630]
119. Kim M, Lima L, Zhao F, Barros J, Medard M, Koetter R, et al. On counteracting Byzantine attacks in network coded peer-to-peer networks. IEEE J Select Areas Commun 2010 Jun;28(5):692-702. [doi: 10.1109/jsac.2010.100607]
120. Chloe A. U.K. High Court orders ISPs to block the pirate bay. PC Magazine. 2012. URL: https://www.pcmag.com/news/297264/u-k-high-court-orders-isps-to-block-the-pirate-bay [accessed 2018-11-05]
121. Wang C, Wang N, Howarth M, Pavlou G. A dynamic peer-to-peer traffic limiting policy for ISP networks. In: Proceedings of the IEEE Network Operations and Management Symposium - NOMS 2010. 2010 Presented at: IEEE Network Operations and Management Symposium - NOMS 2010; April 19-23, 2010; Osaka, Japan URL: https://ieeexplore.ieee.org/document/5488483 [doi: 10.1109/NOMS.2010.5488483]
122. Schulze H, Mochalski K. Internet study 2008/2009. IPoque Report. 2008. URL: https://sites.cs.ucsb.edu/~almeroth/classes/W10.290F/papers/ipoque-internet-study-08-09.pdf [accessed 2018-12-09]
123. Liu Z. Control engineering and information systems. In: Proceedings of the 2014 International Conference on Control Engineering and Information Systems (ICCEIS 2014). 2014 Presented at: International Conference on Control Engineering and Information Systems (ICCEIS 2014); June 20-22, 2014; Yueyang, Hunan, China p. 1-1052. [doi: 10.1201/b17732]
124. Good N, Krekelberg A. Usability and privacy: a study of Kazaa P2P file-sharing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2003 Presented at: CHI03: Human Factors in Computing Systems; April 5 - 10, 2003; Ft. Lauderdale Florida p. 137-144. [doi: 10.1145/642611.642636]
125. Johnson ME. Data hemorrhages in the health-care sector. In: Proceedings of the International Conference on Financial Cryptography and Data Security. 2009 Presented at: International Conference on Financial Cryptography and Data Security; February 23-26, 2009; Accra Beach, Barbados p. 71-89. [doi: 10.1007/978-3-642-03549-4_5]
126. Le FF, Handurukande S, Handurukande AM, Massoulié LA. Clustering in peer-to-peer file sharing workloads. In: Peer-to-Peer Systems III. Berlin: Springer; 2005:217-226.
127. File sharing landscape 2017: where did peer-to-peer network users share which files during 2017-2018. TECXIPIO Magazine. URL: https://www.tecxipio.com/single-post/file-sharing-in-peer-to-peer-networks-2017 [accessed 2018-12-06]
128. Casadesus-Masanell R, Hervas-Drane A. Competing against online sharing. Manag Deci 2010 Sep 07;48(8):1247-1260 [FREE Full text] [doi: 10.1108/00251741011076771]
129. Abdullahi IY, Abdullahi B, Adeshina SA. Towards a framework of configuring and evaluating Modsecurity WAF on Tomcat and Apache web servers. In: Proceedings of the 15th International Conference on Electronics, Computer and

Computation (ICECCO). 2019 Presented at: 15th International Conference on Electronics, Computer and Computation (ICECCO); Dec. 10-12, 2019; Abuja, Nigeria. [doi: 10.1109/icecco48375.2019.9043209]
130. Associated Press. German hospital hacked, patient taken to another city dies. Security Week. 2020. URL: https://www.securityweek.com/german-hospital-hacked-patient-taken-another-city-dies [accessed 2020-12-12]
131. Caporusso N, Chea S, Abukhaled R. A game-theoretical model of ransomware. In: Proceedings of the International Conference on Applied Human Factors and Ergonomics. 2018 Presented at: International Conference on Applied Human Factors and Ergonomics; July 21-25, 2018; Orlando, FL, USA. [doi: 10.1007/978-3-319-94782-2_7]
132. Matthew F. WannaCry cyber-attack cost the NHS £92m as 19,000 appointments cancelled. The Telegraph. 2018. URL: https://www.telegraph.co.uk/technology/2018/10/11/wannacry-cyber-attack-cost-nhs-92m-19000-appointments-cancelled/ [accessed 2019-10-05]
133. Mohurle S, Patil M. A brief study of WannaCry threat: ransomware attack 2017. Int J Adv Res Comput Sci 2017;8(5):4021 [FREE Full text] [doi: 10.26483/ijarcs.v8i5.4021]
134. Ransomware attacks grew by 118%, new ransomware families were detected, and threat actors used innovative techniques. McAfee Labs Threats Report. 2019. URL: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-quarterly-threats-aug-2019.pdf [accessed 2019-10-05]
135. Handler D, Hauge L, Spognardi A, Dragoni N. Security and privacy issues in healthcare monitoring systems: a case study. In: Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies. 2017 Presented at: 10th International Joint Conference on Biomedical Engineering Systems and Technologies; February 21-23, 2017; Porto, Portugal. [doi: 10.5220/0006224603830388]
136. Martin T, Karopoulos G, Hernández-Ramos JL, Kambourakis G, Nai Fovino I. Demystifying COVID-19 digital contact tracing: a survey on frameworks and mobile apps. Wireless Commun Mobile Comput 2020 Oct 17;2020:1-29. [doi: 10.1155/2020/8851429]
137. Zeidler HM. This application is a continuation in part of U.S. application ser. no. 278,001 filed June 24, and now U.S. pat. no. 4,423,287, issued Dec. 27. End-to-end Encryption System and Method of Operation. 1981. URL: https://patents.google.com/patent/US4578530A/en [accessed 2020-07-03]
138. Rösler P, Mainka C, Schwenk J. More is less: on the end-to-end security of group chats in Signal, WhatsApp, Threema. In: Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P). 2018 Presented at: IEEE European Symposium on Security and Privacy (EuroS&P); April 24-26, 2018; London, UK p. 415-429. [doi: 10.1109/eurosp.2018.00036]
139. Cohn-Gordon K, Cremers C, Dowling B, Garratt L, Stebila D. A formal security analysis of the Signal messaging protocol. In: Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P). 2017 Presented at: IEEE European Symposium on Security and Privacy (EuroS&P); April 26-28, 2017; Paris, France. [doi: 10.1109/eurosp.2017.27]
140. Zhao Z, Liu Y, Li H, Yang Y. An efficient user-to-user authentication scheme in peer-to-peer system. In: Proceedings of the First International Conference on Intelligent Networks and Intelligent Systems. 2008 Presented at: First International Conference on Intelligent Networks and Intelligent Systems; Nov. 1-3, 2008; Wuhan, China. [doi: 10.1109/icinis.2008.142]
141. Dehling T, Sunyaev A. Information security and privacy of patient-centered health IT services: what needs to be done? In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: 47th Hawaii International Conference on System Sciences; Jan. 6-9, 2014; Waikoloa, HI, USA p. 2984-2993. [doi: 10.1109/hicss.2014.371]
142. Levine BN, Shields C, Margolin NB. A survey of solutions to the sybil attack. University of Massachusetts Amherst, Amherst, MA. 2006. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.6395&rep=rep1&type=pdf [accessed 2018-10-05]
143. Damiani E, di Vimercati DC, Paraboschi S, Samarati P, Violante F. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In: Proceedings of the 9th ACM Conference on Computer and Communications Security. 2002 Presented at: CCS02: ACM Conference on Computer and Communications Security; November 18 - 22, 2002; Washington, DC USA p. 207-216. [doi: 10.1145/586110.586138]
144. van Vroonhoven J. Peer to peer security. In: Proceedings of the 4th Twente Student Conference on IT. 2006 Presented at: 4th Twente Student Conference on IT; January 30, 2006; Enschede, The Netherlands p. 1-10 URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=D4C9D04083126B780B5BB6CC02E129C7?doi=10.1.1.60.3334&rep=rep1&type=pdf
145. Chaokai H, Meng W. Comparison and analysis of different reputation systems for peer-to-peer networks. In: Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE). 2010 Presented at: Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE); Aug. 20-22, 2010; Chengdu, China p. 20-23. [doi: 10.1109/icacte.2010.5579838]
146. Banik SM, Pena L. Deploying agents in the network to detect intrusions. In: Proceedings of the IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS). 2015 Presented at: IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS); June 28 - July 1, 2015; Las Vegas, NV, USA p. 83-87. [doi: 10.1109/icis.2015.7166574]
147. Sharma O, Girolami M, Sventek J. Detecting worm variants using machine learning. In: Proceedings of the International Symposium on Electronic Commerce and Security. 2008 Presented at: ACM Conference on Emerging Network Experiment

and Technology, CoNEXT 2007; December 10-13, 2007; New York, USA p. 1-12 URL: https://doi.org/10.1145/1364654.1364657 [doi: 10.1145/1364654.1364657]
148. Jain C, Saxena AK. General study of mobile agent based Intrusion Detection System (IDS). J Comput Commun 2016;04(04):93-98 [FREE Full text] [doi: 10.4236/jcc.2016.44008]
149. Shah B, Trivedi BH. Improving performance of mobile agent-based intrusion detection system. In: Proceedings of the Fifth International Conference on Advanced Computing & Communication Technologies. 2015 Presented at: Fifth International Conference on Advanced Computing & Communication Technologies; Feb. 21-22, 2015; Haryana, India. [doi: 10.1109/acct.2015.118]
150. Zhou CV, Karunasekera S, Leckie C. A peer-to-peer collaborative intrusion detection system. In: Proceedings of the 13th IEEE International Conference on Networks Jointly held with the 2005 IEEE 7th Malaysia International Conference on Communications. 2005 Presented at: 13th IEEE International Conference on Networks Jointly held with the 2005 IEEE 7th Malaysia International Conference on Communications; Nov. 16-18, 2005; Kuala Lumpur, Malaysia. [doi: 10.1109/icon.2005.1635451]
151. Hosszú G, Czirkos Z. Network-based intrusion detection. In: Encyclopedia of Internet Technologies and Applications. Hershey, Pennsylvania: IGI Global; 2008:353-359.
152. Graham S, Fu X, Lu B. Network security fundamentals. In: Wiley Encyclopedia of Computer Science and Engineering. Hoboken, New Jersey, United States: John Wiley & Sons; 2008.
153. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med 2000 Oct;51(7):1087-1110. [doi: 10.1016/s0277-9536(00)00098-8] [Medline: 11005395]
154. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
155. Morana M. How to start a software security initiative within your organization: a maturity based, and metrics driven approach. The OWASP Foundation. 2009. URL: https://www.owasp.org/images/c/c4/OWASP-ItalyDayEGov09_04_Morana.pdf [accessed 2018-11-02]

Abbreviations

CIA: confidentiality, integrity, availability
CVSS: Common Vulnerability Scoring System
DDoS: distributed denial-of-service
DHT: Distributed Hash Tables
DLT: distributed ledger technology
DoS: denial-of-service
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
HTI: health care technology infrastructure
IAS: identity authentication schemes
IDS: intrusion detection systems
IHE: integrating health care enterprise
ISP: internet service provider
MitM: man-in-the-middle
P2P: peer-to-peer
PEPP-PT: Pan-European Privacy-Preserving-Proximity-Tracing
PHS: patient-centered health care information system
PRS: pure random scan
TCP: transmission control protocol
TRM: trust and reputation model

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by H Kondylakis, C Friedrich, C Fitte, F Teuteberg; comments to author 26.11.20; revised version received 20.05.21; accepted 02.08.21; published 15.11.21.

Please cite as:
Abdullahi Yari I, Dehling T, Kluge F, Geck J, Sunyaev A, Eskofier B
Security Engineering of Patient-Centered Health Care Information Systems in Peer-to-Peer Environments: Systematic Review
J Med Internet Res 2021;23(11):e24460
URL: https://www.jmir.org/2021/11/e24460
doi:10.2196/24460
PMID:34779788

Review

Dangers and Benefits of Social Media on E-Professionalism of Health Care Professionals: Scoping Review

Tea Vukušić Rukavina[1], MD, PhD; Joško Viskić[2], DMD, PhD; Lovela Machala Poplašen[1], MLiS; Danko Relić[1], MD, PhD; Marko Marelić[1], MSc; Drazen Jokic[3], DMD; Kristijan Sedak[4], MagComm

[1]Andrija Stampar School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia

[2]Department of Fixed Prosthodontics, School of Dental Medicine, University of Zagreb, Zagreb, Croatia

[3]Department of Orthodontics, School of Dental Medicine, University of Zagreb, Zagreb, Croatia

[4]Department of Communication Sciences, Catholic University of Croatia, Zagreb, Croatia

Corresponding Author:
Joško Viskić, DMD, PhD
Department of Fixed Prosthodontics
School of Dental Medicine
University of Zagreb
Gundulićeva 5
Zagreb, 10000
Croatia
Phone: 385 98769245
Email: viskic@sfzg.hr

Abstract

Background: As we are witnessing the evolution of social media (SM) use worldwide among the general population, the popularity of SM has also been embraced by health care professionals (HCPs). In the context of SM evolution and exponential growth of users, this scoping review summarizes recent findings of the e-professionalism of HCPs.

Objective: The purpose of this scoping review is to characterize the recent original peer-reviewed research studies published between November 1, 2014, to December 31, 2020, on e-professionalism of HCPs; to assess the quality of the methodologies and approaches used; to explore the impact of SM on e-professionalism of HCPs; to recognize the benefits and dangers of SM; and to provide insights to guide future research in this area.

Methods: A search of the literature published from November 1, 2014, to December 31, 2020, was performed in January 2021 using 3 databases (PubMed, CINAHL, and Scopus). The searches were conducted using the following defined search terms: “professionalism” AND “social media” OR “social networks” OR “Internet” OR “Facebook” OR “Twitter” OR “Instagram” OR “TikTok.” The search strategy was limited to studies published in English. This scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines.

Results: Of the 1632 retrieved papers, a total of 88 studies were finally included in this review. Overall, the quality of the studies was satisfactory. Participants in the reviewed studies were from diverse health care professions. Medical health professionals were involved in about three-quarters of the studies. Three key benefits of SM on e-professionalism of HCPs were identified: (1) professional networking and collaboration, (2) professional education and training, and (3) patient education and health promotion. For the selected studies, there were five recognized dangers of SM on e-professionalism of HCPs: (1) loosening accountability, (2) compromising confidentiality, (3) blurred professional boundaries, (4) depiction of unprofessional behavior, and (5) legal issues and disciplinary consequences. This scoping review also recognizes recommendations for changes in educational curricula regarding e-professionalism as opportunities for improvement and barriers that influence HCPs use of SM in the context of e-professionalism.

Conclusions: Findings in the reviewed studies indicate the existence of both benefits and dangers of SM on e-professionalism of HCPs. Even though there are some barriers recognized, this review has highlighted existing recommendations for including e-professionalism in the educational curricula of HCPs. Based on all evidence provided, this review provided new insights and guides for future research on this area. There is a clear need for robust research to investigate new emerging SM platforms, the efficiency of guidelines and educational interventions, and the specifics of each profession regarding their SM potential and use.

(J Med Internet Res 2021;23(11):e25770) doi:10.2196/25770

KEYWORDS

e-professionalism; social media; internet; health care professionals; physicians; nurses; students; medicine; dental medicine; nursing

Introduction

Background

Global digital growth shows no sign of slowing, with a million new people worldwide coming online every day. This growth is clearly fueling social media (SM) use, as 45% of the world's population are now SM users: a whopping 3.5 billion people [1]. The popularity and use of SM has increased substantially in the past few years, despite controversy around privacy, hacking, fake news, and all other negative aspects of online life [2].

Social media have been defined as "a group of online applications that allow for the creation and exchange of content generated by users" [3] and can be categorized into five groups: (1) collaborative projects (eg, Wikipedia), (2) blogs or microblogs (eg, Blogger or Twitter), (3) content communities (eg, YouTube), (4) social networking sites (SNSs; eg, Facebook), and (5) virtual gaming or social worlds (eg, Second Life) [4]. SNSs (eg, Facebook) are "applications that enable users to connect by creating personal information profiles, inviting friends and colleagues to have access to those profiles, and sending e-mails and instant messages between each other" [4]. There is a lot of mixing and confusion between the terms SM and SNSs with SM being a newer and a much broader term, encompassing SNSs.

As we are witnessing the evolution of SM use globally among the general population, popularity of SM has also been embraced by health care professionals (HCPs) [5]. It is further reflected as a considerable growth in the research about SM use in health and medicine [6], mainly focusing on the roles of SM or SNSs in linking patients and HCPs [7-9] or use of SM/SNSs for communication among HCPs [10].

Within these new platforms exists an unprecedented ability to expand access and communication, with the potential to revolutionize the way medical professionals interact with peers, patients, and the public. However, along with this expanded access lies the potential for inadvertent overlap between the physicians' personal and professional lives. Boundary concerns are increasing with the blurring of personal and professional lines on SM [11]. Anything placed on the internet is essentially permanent, and our "digital footprint" stays forever documented in this virtual yet, for almost everyone, accessible world.

A definition for a new term "e-professionalism" was given by Cain and Romanelli [12] as "attitudes and behaviors (some of which may occur in private settings) reflecting traditional professionalism paradigms that are manifested through digital media." The intersection between medical professionalism and SM has been termed also as online professionalism or digital professionalism [13].

Prior reviews have focused on the e-professionalism of medical students, residents, or physicians [14], or have presented a full spectrum of SM-related challenges and opportunities in the context of medical professionalism of diverse types of HCPs [15,16], but these studies were conducted almost 6 years ago. Within that time frame, the number of SM users (both HCPs and patients), SM influence on private and professional life, and new features within SM have increased substantially leaving scientific research struggling to keep up [17-19]. Since the end of November 2014, the number of Facebook users has doubled (from 1.35 billion to 2.7 billion) and the number of Instagram users almost quadrupled (from 300 million to 1.158 billion). New SM platforms like TikTok have gained popularity [20]. A gap has emerged in comprehensive understanding of the ways SM has influenced the medical field, especially professional behavior. With time being crucial in the context of SM evolution and exponential growth of users, this scoping review maps and summarizes recent findings and fills the knowledge gap about e-professionalism of HCPs.

Objective

The purpose of this scoping literature review was to characterize the recent original peer-reviewed research studies published between November 1, 2014, to December 31, 2020, on e-professionalism of HCPs; to assess quality of the methodologies and approaches used; to explore the impact of SM on e-professionalism of HCPs; to recognize benefits and dangers of SM; and to provide insights to guide future research in this area.

Methods

A Scoping Review

We performed a scoping review to explore the extent of the latest current evidence on e-professionalism of HCPs. The review questions were what are the reported outcomes of the benefits and dangers, and of SM on e-professionalism for HCPs, and what was the quality of methodologies and approaches used on the use of SM affecting e-professionalism of HCPs? The review subquestions were which knowledge gaps have been identified in studies and what are the recommendations for future research?

The scoping review method was chosen to map and summarize the evidence, and inform future research in the domain of e-professionalism of HCPs [21-23]. We performed a scoping review consistent with the guidance provided by the Joanna Briggs Institute Reviewer Manual [24,25]. The scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [26]. The protocol was registered on the Open Science Framework (registration DOI 10.17605/OSF.IO/YR8TW) [27] on April 9, 2021.

Search Strategy

A search of the literature was performed in January 2021 using 3 databases (PubMed, CINAHL [EBSCO], and Scopus). The searches were conducted for the period from November 1, 2014, to December 31, 2020, using the following defined search terms: "professionalism" (a Medical Subject Headings [MeSH] term) AND ("social media" [a MeSH term] OR "social networks" OR "Internet" [a MeSH term] OR "Facebook" OR "Twitter" OR "Instagram" OR "TikTok") included in the title or abstract or keywords. The search strategy was limited to studies published in English.

The full search strategy is summarized in Multimedia Appendix 1. A senior information specialist validated the search strategy. For a comprehensive assessment, we also searched the reference lists of all the included articles to identify other studies that may be relevant to our review.

Study Inclusion and Exclusion Criteria

Studies were included in this review if they were original research focused primarily on the use of SM among health professionals and studied the uses, benefits, or dangers of SM.

Studies were excluded from this review if they were not in English; were books, dissertations, reviews, reports, abstracts only, case studies, opinions, letters, commentaries, policies, guidelines, or recommendations; did not focus primarily on the use of SM among health professionals; did not study the uses, benefits, or dangers of SM among health professionals; did not study HCPs as a study population; described the use of SM primarily with a marketing or advertising focus; and were not available as full text in the final search.

Eligibility and Data Extraction

Following the search, all references captured by the search engine were uploaded to the reference management system Zotero (Corporation for Digital Scholarship). Duplicates were identified and removed. Initial screening of the studies, based on the information contained in the titles and abstracts, was undertaken independently by 2 reviewers (TVR and JV). The interrater reliability (IRR) between them was established. IRR to screen the papers was determined using the indices *average pairwise percent agreement*, *Cohen kappa*, and *Krippendorff* α (alpha) [28]. IRR was calculated with the ReCal ("Reliability Calculator"), an online utility that computes IRR coefficients [29]. To assess the eligibility of the papers, two researchers (TVR and JV) independently reviewed and evaluated the studies. Discrepancies were discussed with reference to the research objectives until consensus was reached on the inclusion for the analysis.

Data extraction was done in two passes. In the first pass, seven reviewers (TVR, JV, DR, LMP, DJ, KS, MM) extracted the data from included studies; in the second pass, two reviewers (TVR and JV) agreed on the extracted data. The following data were extracted: authors, year, country of origin, study objective and design, data analysis methods, study population, type of SM, key measures, conclusions/recommendations, and ethics statements.

Quality Evidence Assessment

The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies developed by the National Heart, Lung, and Blood Institute was used to evaluate the quality of the quantitative studies [30]. The Critical Appraisal Skills Programme appraisal tools were used to evaluate the quality of the qualitative and mixed methods reviewed studies [31]. Quality evidence assessment was performed by two reviewers (TVR and JV). If there was a discrepancy, a third reviewer was consulted (LMP).

Ethics Approval and Consent to Participate

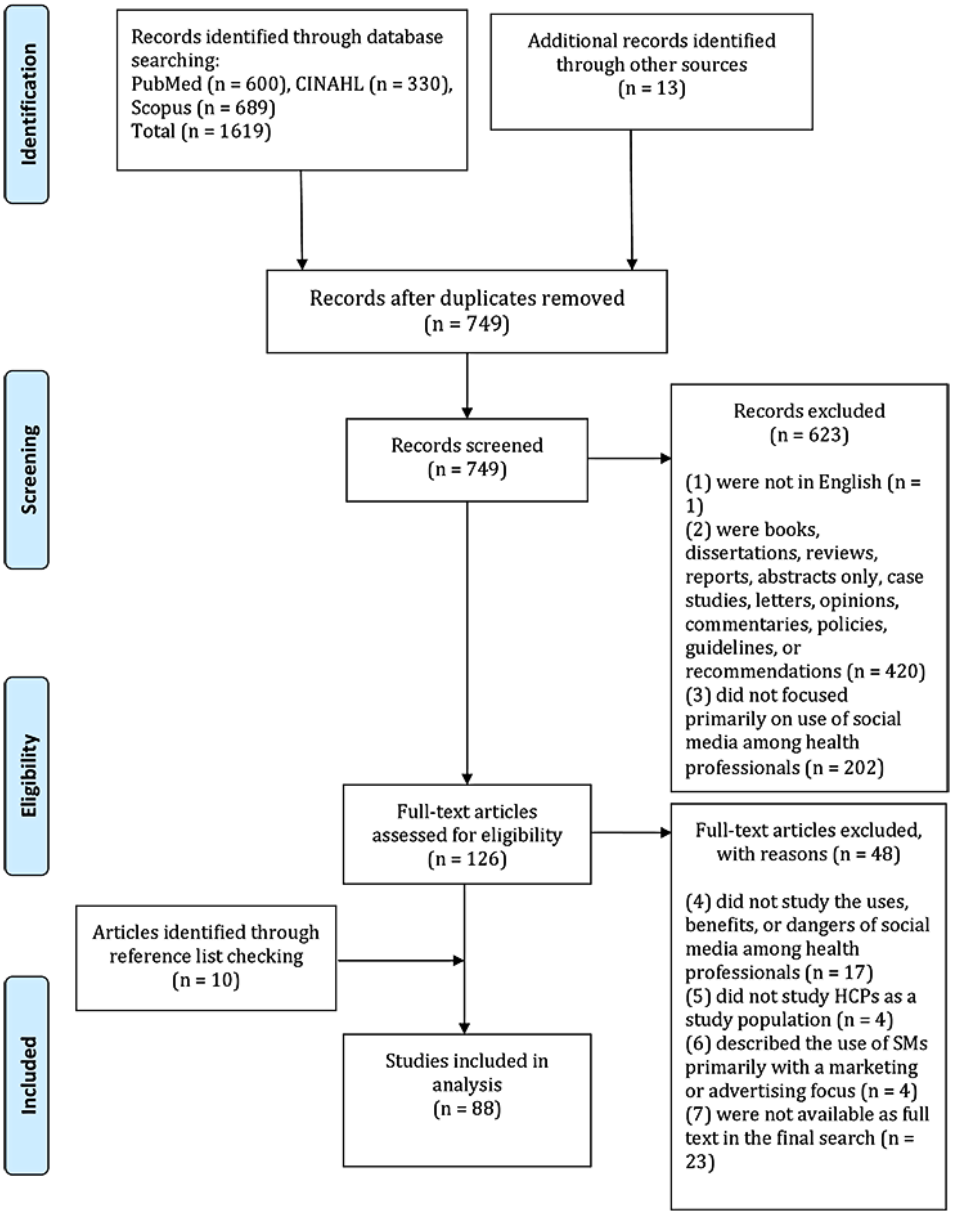
This was a scoping review with no data collected from human participants. Ethical approval was not needed.

Results

Findings

Figure 1 shows the PRISMA flowchart illustrating the searching process and how the studies were included in the review. The literature search retrieved 1632 papers, and after removing duplicates, 749 titles and abstracts were screened. The IRR between the two researchers (TVR and JV) who screened titles and abstracts was established. IRR values indicated high reliability (average pairwise percent agreement 89%, Cohen kappa 0.80, Krippendorff α=.83). Full texts of 126 papers were assessed for eligibility. We also searched the reference lists of the included articles and found another 10 relevant articles for inclusion. Thus, a total of 88 studies were finally included in this review. Details of the studies including year, country of origin, study objective and design, data analysis methods, study population, type of SM, key measures, conclusions/recommendations, and ethics statements are shown in Multimedia Appendix 2. The studies (n=671) that were excluded are shown in Multimedia Appendix 3, along with the reasons for their exclusion.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection procedure. HCP: health care professional; SM: social media.



Characteristics of the Reviewed Studies

The studies were conducted in 40 countries, the majority being based in the United States (n=35), Canada (n=13), Australia (n=10), and the United Kingdom (n=10). Other countries with one or two studies were Brazil, China, Greece, Dominican Republic, Germany, Hong Kong, India, Ireland, Italy, Kingdom of Saudi Arabia, Mexico, New Zealand, Nigeria, Pakistan, Philippines, Singapore, South Africa, South Korea, Spain, Taiwan, Thailand, Turkey, and the United Arab Emirates (Multimedia Appendix 2). A total of 8 studies were conducted in several countries simultaneously [32-39]. Participants in the reviewed studies were from diverse health care professions (Table 1).

Medical health professionals were involved in about three-quarters of the studies. On several occasions, more than one health care profession was involved in the evaluated studies. Regarding the educational level of targeted HCPs, 35 studies investigated students [32,40-73]; 41 studies investigated residents or practicing HCPs [34-39,74-108]; 5 studies investigated deans or directors of programs [109-113]; and 7 studies investigated several educational levels of HCPs, students, residents, faculty members, or practicing HCPs [33,114-119].

Table 2 describes the types of SM/SNSs studied; the majority of the studies were unspecific, studying use of any type of SM or SNS. Only Facebook or "all SMs/SNSs with specific reference to Facebook" was analyzed in one-third of the studies (Multimedia Appendix 2). Twitter [38,44,80,91,110], Instagram [101], and YouTube [37] were specifically targeted SM/SNSs in 7 studies.

Table 1. Types of health care professions included in the reviewed studies.

Health care profession	Studies (N=88), n[a]
Medical	
Deans and directors	2
Faculty	6
Specialists	11
Doctors (general)	8
Residents	15
Students	23
Dental	
Program directors	2
Faculty	1
Doctors (general)	1
Students	5
Nursing	
Deans and directors	1
Faculty	2
Nurses	3
Students	8
Pharmacy	
Pharmacists	2
Students	4
Other health care professionals[b]	4

[a]A study could include more than one type of health care professional.

[b]Other health care professionals are physiotherapists, physician assistant students, and osteopathic medicine students.

Table 2. Types of social media or social networking sites.

Social media/social networking site	Studies, n
Unspecific (any type of social media/social networking sites)	59
Facebook	21
All social media sites with specific reference to Facebook	1
All social media sites with specific reference to Facebook and Twitter	3
Instagram	1
Twitter	2
YouTube	1

Assessing the Quality of Studies

Overall, the quality of the studies was satisfactory. Most of the reviewed studies met the criteria in checklists (Multimedia Appendix 4). All studies were exploratory in nature, and the findings were descriptive. Among 88 studies, 49 were quantitative [32, 33, 35, 38-43, 51-53, 55, 59-61, 64, 66-69, 71, 72, 74, 75, 77-80, 82, 85, 88, 89, 91, 94-96, 98, 100, 102, 105-108, 110, 113, 114, 116, 118], 12 were qualitative [34, 36, 44, 45, 49, 57, 73, 76, 92, 93, 101, 104], and 27 used mixed methods [35, 37, 46-48, 50, 54, 56, 58, 62, 63, 65, 70, 81, 83, 84, 86, 87, 90, 97, 99, 103, 109, 111, 112, 115, 117]. Most studies used surveys (n=64) [32, 33, 35, 38-43, 47, 48, 51-55, 57, 59-62, 64-72, 74, 75, 77-82, 85, 88, 89, 91, 94-98, 100, 102, 103, 106-119]. The questionnaires used in surveys were mostly developed by researchers. Of these survey studies, only about one-third had a response rate of 50% or greater, and 9 studies did not explicitly report a response rate [39,53,60,63,65,67,72,75,96]. In the mixed methods studies, dominantly, content or thematic analyses were used. A total of 11 studies conducted in-depth or semistructured interviews [34,36,44,49,58,65,73,76,97,104,117], and 6 studies used focus

groups [45,49,58,70,112,117]. Most studies included (n=77) had clear ethical statements within the paper either stating ethical board approval or exemption, and 11 studies did not explicitly report an ethical statement [37-39,53,54,72,74,90,105,108,117].

SM Use Patterns Among Health Care Professionals

Studies that assessed SM use among different types of HCPs found high use among students, from 66.9% to 98.7% [40,43,50,51,60,61,64-66,68-70,102]; the highest with 98.7% using Facebook at least once a week was established among dental students. Lower rates of use were seen in practicing HCPs, physicians of different specialties, and program directors (PDs) or faculty, mostly ranging from 50% to 80% [35,39,74,80,83,85,86,88,89,91,96,105,106,108,110,113,114]. The exceptions were 3 studies: family medicine residents and physicians in Saudi Arabia where 95.4% of participants reported having an SM account and checking them at least once a day [79], 93.4% of medical doctors in a Singapore hospital [100], and 100% of Chinese registered nurses owned an SM account [107].

Several studies demonstrated a "generation gap" in SM use, where students are more likely than faculty to use SM [114,115,117,118]. A linear relationship between increasing age and decreasing SM use was also found among physicians of the same specialty or other HCPs [85,91,94,95,100,106].

Significant gender differences were established in several studies [71,79,91,95,105,107].

Irfan et al's [79] study showed females using SM more for professional purposes, and Wang et al's [107] study, where the study population was registered nurses, was similar. In Patel et al's [91] study where a subgroup analysis on Twitter use for professional purposes revealed a significant gender difference: only one in four users was a female radiologist and only 14% of highly active users were female. Gender difference was established also in Gupta et al's [62] content study of Facebook profiles (in favor of male medical students, ranging from 73.5% to 96.4%) [62], and 98.8% of all participants were males in a study about orthopedic surgeons [105].

Studies showed privacy settings deployment from 71% to 97% among HCPs [40,42,65,66,68,96,100,115,117,118]. Only 4 studies explicitly stated the percentage of students who had set their Facebook account on private: 37% of pharmacy students [40], 83.6% and 91% of dental students [64,65], and 71% of medical students [42]. More students than faculty used privacy features [115,118]. Results from a study among doctors in Singapore suggest that there is a knowledge deficit in terms of understanding of the privacy settings of SM accounts. From 30% to 55% of the respondents had an incorrect understanding of their SM account settings despite 95.5% claiming that they were aware that the institution had a SM policy [100]. Use of real names was investigated in 2 studies; in both, the vast majority of HCPs used their own names on Facebook [51,117], but on Twitter (45%) and Instagram (54%), far fewer dental students presented themselves with real names [51].

Studies also investigated the purpose of SM use, whether participants mixed professional and personal information and activities on SM sites (blended profiles) or adopted a separation strategy where professional information and activities were clearly separated from personal ones (dual citizenship) [34,39,69,78,79,91,96,97,106]. In Duke et al's [115] study, significant differences were established between nursing students' and faculties' purpose of use, where almost twice as many students used SM for educational purposes than did faculty (58.5% vs 27.6%; P<.001), and almost 96% of students used SM to talk about academic-related problems compared to only 28% of faculty who did so (P<.001). Irfan et al [79] investigated family medicine residents and physicians use of SM not only for personal purposes (76%) or professional reasons (26%); they have determined that participants also use SM for general education (46%) and in a smaller proportion (18.2%) for continuing medical education [79]. A study among South African nurses established that they experience difficulty in separating their professional and personal lives when using SM [73].

In terms of themes covered in the reviewed studies, we have focused on two major themes: benefits and dangers of SM on e-professionalism of HCPs. This scoping review also recognizes an evaluation of effects of existing approaches on promoting e-professionalism and barriers that influence HCPs use of SM in the context of e-professionalism.

Benefits of Social Media on E-Professionalism of Health Care Professionals

For the selected studies, there are three recognized benefits of SM on e-professionalism of HCPs: (1) professional networking and collaboration, (2) professional education and training, and (3) patient education and health promotion.

Professional Networking and Collaboration

The benefits of SM on e-professionalism of HCPs can be seen as improvements of established networks or possibilities for collaboration through SM sites [33, 34, 36, 38, 39, 44-46, 53, 69, 72, 76, 78, 79, 92, 97, 106-108, 117]. Besides providing the opportunity for connecting with others and sharing experiences [38,39,46,53,69,72,78,106-108], SM have enabled the creation of communities for support. This enables students to help each other in studying and interacting with faculty who can provide advice, encouragement, and virtual mentorship [44]. Chretien et al [44] describe two roles that medical students use via SM, first, as access to information and, second, as a voice, a platform for advocacy, and an opportunity to state attitudes and opinions. SM is also where they gained control of their digital footprint and a sense of equalization within the medical hierarchy. SM provide for students, residents, and faculty a good discussion medium and an engaging way to get high-quality current information [79,117]. Professional networking and collaborations on SM enable the development and building of professional identities for health care professions [34,38,39,45,67,73,104]. Some studies emphasize SM benefits of peer-to-peer advising or learning, provision of emotional support, and identifying approaches through which physicians establish interpersonal trust on SM [36,45,92].

Professional Education and Training

Several studies have demonstrated students' use of SM for acquiring knowledge, to gain access to information from experts

with whom they otherwise would not be able to connect with, or for creating communities that can then be used as means for supportive, professional, and social learning [44,45,58,66].

A survey among US surgeons indicated that 70% of respondents believe SM benefits professional development, similar to findings among Chinese urologist where 52.7% believe that SM provides a platform for surgical or medical education [95,106].

Canadian urologists (59%) consider SM as a simple information repository that is likely to increase in the continuing professional development space [95]. About 80% of HCPs from Saudi Arabia agreed with the benefits of using SM in health care services and considered that the use of these technologies in the provision of health services improves their professional knowledge and that SM can be a useful tool by which physicians may promote their services [75].

In Duke et al's [115] study, significant differences were established between nursing students' and faculties' purpose of use, where almost twice as many students used SM for educational purposes than did faculty (58.5% vs 27.6%; P<.001), and almost 96% of students used SM to talk about academic-related problems compared to only 28% of faculty who did so (P<.001). Irfan et al [79] investigated family medicine residents' and physicians' use of SM for not only personal purposes (76.0%) or professional reasons (26.0%); they have determined that participants use SM also for general education (46.0%) and, in a smaller proportion, for continuing medical education (18.2%).

Patient Education and Health Promotion

Positive professional behaviors and attitudes regarding patient education and health promotion were also reported [34,37,75,97,106]. George et al [47] investigated US medical students' attitudes about what positive role SM can play in improving communication with patients. A total of 44% of respondents stated they should and would react if a patient sought their medical advice via Facebook. Some students acknowledged the potential usefulness of SM in medical practice, patient education, health promotion, and interpersonal communication, if applied in a safe and responsible manner [47]. The thematic analysis of pharmacists' semistructured interviews recognized addressing unprofessional posts made by peers as positive online behavior. Another positive professional activity was the use of SM to educate society in general about the role that pharmacists play in the health care system, their clinical roles, and how they can promote quality care for patients [34]. More than half of the HCPs in a cross-sectional study in Saudi Arabia agreed with the benefits of using SNSs in health care services as a suitable tool for patient education and raising public health awareness [75]. The results of the study among Saudi Arabian orthopedic surgeons showed that they are more likely to post online for the sake of sharing general medical knowledge as opposed to giving specific treatment advice. Most of them were open to the possibilities of using SM more with their patients for the sake of education, knowledge sharing, and improving patient outcomes [97]. In terms of communication with patients, in Long et al's [106] study, the majority of urologists thought SM had improved efficiency in patient education (65.4%) and patient communication (55.1%).

Dangers of Social Media on E-Professionalism of Health Care Professionals

For the selected studies, there are five recognized dangers of SM on e-professionalism of HCPs: (1) loosening accountability, (2) compromising confidentiality, (3) blurred professional boundaries, (4) depiction of unprofessional behavior, and (5) legal issues.

Loosening Accountability

According to some studies in this review, loosening accountability can be seen as a danger to e-professionalism from two points of view: eroding public trust by providing poor quality of information on SM [39,106,117] and damaging to the professional image [43, 45, 51, 56, 57, 59, 64, 66, 68, 70, 73, 102, 106, 112].

Potential damage to the professional image has been depicted by students as concerns about repercussions of their posts on career development or future employment, since employers are checking SM profiles of candidates [43, 45, 51, 56, 57, 59, 66, 68, 70].

Students are concerned about the extent of representation of the students' character on SM; they edit profiles before interviews or career fairs [57,70] or intend to review or modify their profiles when they become qualified [45,51]. As students get closer to graduation, they are more concerned about future employment opportunities and their professional career. In addition, it has been reported that there is more awareness of online responsibilities as students progress through their program because employers can, and at times do, use SM profiles to make hiring decisions [56].

Three reviewed studies investigated PDs' (medical and dental) attitudes about the use of SM for admission criteria [109,110,112]. In a study that evaluated how SM is being used in dental hygiene programs admissions and policy, only 4% of programs evaluated a potential student's internet presence, mostly by searching on Facebook. Of those respondents that do not evaluate internet presence in applicants, more than half are not considering adding this to the admissions criteria (57.2%). Others are considering it (39.1%), and a small number (3.6%) plan to implement this in the future [109]. Use of SM is higher among medical PDs, and they more often view the online behavior of residency applicants, surgical residents, and faculty surgeons [110]. Among general surgery PDs, 18% reported visiting the SM profiles of medical students applying for surgical residency. Overall, 11% of PDs reported lowering the rank or completely removing a residency applicant from the rank order list because of online behavior [110].

Compromising Confidentiality

Being both an ethical and potentially legal issue, many of the studies have investigated attitudes toward compromising confidentiality, concerns that HCPs have about use of SM, patient privacy, and violations of Health Insurance Portability and Accountability Act (HIPAA) standards as a separate problem. Breaches of patient privacy was a concern for many

different types of HCPs [34,43,47,57,91,94,97,107,111,119]. Bagley et al's [41] results showed that the frequency of a student's updates of a Facebook status appears to be associated with a risk of violating HIPAA online. Similar findings were made in Wejis et al's [96] study. Greater disclosure on Facebook was associated with lesser awareness of the consequences of posting information on Facebook, a greater need for popularity, a higher level of self-esteem, a greater number of Facebook friends, and a higher frequency of signing in to Facebook [96]. In a study among nursing students, perceptions of confidentiality existed on the level of knowledge; all students knew that posting patient names or pictures was a breach of confidentiality. However, 34% were aware of other students who had breached patient confidentiality on Facebook [43]. "Cognitive dissonance," a disconnect between what they thought they *would* do versus what they thought they *should* do was also reported by George et al [47].

Examples of compromising confidentiality and breaches of patient privacy were reported in several studies. In Long et al's [106] study among Chinese urologists, nearly half of the respondents had experience posting information or pictures of patients' SM, but only 5% of them sought their patients' consent before posting [106]. In an exploratory qualitative study among nursing students in South Africa, students admitted that there is no responsible use of SM. They have stated that each of them perceives responsible use of SM differently. They took pictures, recorded video and audio clips of patients and of clinical interactions involving patients, and posted this information on SM compromising confidentiality [73]. In Wang et al's [107] study, 13.4% of Chinese registered nurses (n=88) confessed that they had "sometimes" posted anonymous patient information on SM.

Blurred Professional Boundaries

Traditional boundaries are blurred on many levels by online interactions. Blurred boundaries between professional and personal spheres of SM use [34, 39, 47, 51, 69, 78, 79, 91, 96, 97, 106], with concerns about exposure of one's private life or separating private and professional profiles have been presented in numerous studies of this review.

Several studies in this review investigated the purpose of SM use and whether participants mixed professional and personal information and activities on SM sites (blended profiles) or adopted a separation strategy where professional information and activities were clearly separated from personal ones (dual citizenship) [34,78,79,91,96].

Numerous studies document blurred boundaries between patient and HCPs, and between students and faculty [40, 43, 47, 59, 64, 65, 68, 70, 73, 75, 78, 82, 85, 98, 100, 105, 107, 114, 117]. Medical students have different attitudes regarding online interaction with a patient. "Friending a patient" is generally not acceptable nor endorsed; a wide range of opinions have been observed concerning this issue, ranging from one-third for medical students in Brazil [59] that find this unacceptable to 92% for senior medical students in New Zealand [70].

Among physicians, the majority have legal concerns about communicating with patients through SM [78,105,117]. In Fuoco and Leveridge's [78] study about attitudes toward and use of SM among urologists, online patient interaction was endorsed by only 14% of urologists. Even though 56% of urologists agreed that SM integration in medical practice will be "impossible" due to privacy and boundary issues, 73% felt that online interaction with patients would become unavoidable in the future, especially for those in practice [78].

Students were anxious about the possibility their teachers could read about their personal life on SM. Dental students are ambivalent toward "friending" a faculty member [65]. From pharmacy students' perspectives, an active user is generally open to "friending" the outside world. However, the majority were still reluctant to "friend" faculty members at their school. Students have beliefs that student-faculty interactions should remain professional, and SM sites are not appropriate venues for such professional communication [40]. Academic faculty members were worried that connecting via SM with students or residents would blur the boundaries of the teacher-student relationship. In Jafarey et al's [117] study, almost half of faculty members found it inappropriate to friend a current student, and friending patients was not acceptable for 70% of respondents, with major differences found in age groups; it was acceptable to friend patients to 31% of trainees and 62% of students compared to only 5% of faculty. In a similar linear progression, younger age associated with more openness to being friends with patients was also demonstrated by Klee et al [82]. Two-thirds of family medicine residents and half of practicing physician respondents believed it was not ethical to be SM friends with patients.

Brisson et al [114] found that faculty were more likely than students to have been approached by patients on SNSs (53% vs 3%). Karveleas et al's [64] study showed a significantly higher percentage of fifth-year dental students (48.3%) compared to fourth-year students (20.6%; $P<.001$) who had received a Facebook friend request from one or more patients [64].

Depiction of Unprofessional Behavior

Numerous studies in this review have tried to assess the extent of unprofessional behavior posted by HCPs themselves or seen to be posted by their peers. Although there is no uniform consensus on what constitutes unprofessional behavior, studies most frequently associated it with online content pertaining to alcohol intoxication; substance or illegal drug use, nudity, and sexuality; demeaning content about patients, peers, educators, clinical sites, or the profession as a whole; discriminatory content; profanity; and aggressive/bullying content toward coworkers. Surveys that captured students' self-report of posted unprofessional behavior reported witnessing the investigated examples with varying frequencies [32, 42, 43, 55, 59, 62, 64, 73, 114, 118]. Among Brazilian medical students, frequencies ranged from 13.7% for "violation of patient's privacy" to 85.4 % for "photos depicting consumption of alcoholic beverages" [59]. Posting of unprofessional content was highly prevalent among medical students in Australia despite understanding that this might be considered inappropriate and despite awareness of professionalism guidelines. A total of 34.7% of students reported unprofessional content (eg, evidence of being intoxicated 34.2%, illegal drug use 1.6%, posting patient

information 1.6%, and depictions of an illegal act 1.1%) [36]. In Kenny and Johnson's [51] study among dental students, 34% had questionable content on their profile, while 3% had definite violations of professionalism on their profile and 25% had unprofessional photographs on their profile including alcohol and different levels of nudity. Of those with unprofessional photographs, 52% had a documented affiliation with the dental school also visible on their profile [56]. In another study among dental students by Karveleas et al [64], unprofessional content had been posted by most students. A total of 71.7% of students had posted pictures from holidays, 41.5% moments in nightclubs, and 26.2% photographs wearing swimwear or underwear. Alcohol consumption and smoking were published by 19.1% and 5.5% of responders, respectively, while 0.4% of responders admitted having posted photographs of themselves using illegal drugs [64].

An international survey among health science students, from 8 universities in 7 countries, registered that a significant number of students (20.5%) across all health science disciplines self-reported sharing clinical images inappropriately [32]. Furthermore, medical students who observed unprofessional behaviors were more likely to participate in such behaviors [55], and the phenomenon of "distancing" was described among nursing students, while the existence of unwise posting on SNSs was widely acknowledged, students tended to attribute such behavior to others [43].

Age difference in the terms "older and wiser," meaning more cautious about posting unprofessional behavior online, was proven in studies comparing students' and faculties' online behavior. Medical students were more likely than faculty to display content they would not want patients to see (57% vs 27%), report seeing inappropriate content on colleagues' SNS profiles (64% vs 42%), and ignore harmful postings by colleagues (25% vs 7%) [114]. Medical students in Kitsis et al's [118] study reported the self-posting of profanity, depiction of intoxication, and sexually suggestive material more often than faculty ($P<.001$). Medical students and faculty both reported peers posting unprofessional content significantly more often than self-posting [118].

Studies that assessed the online unprofessional behavior of residents or practicing HCPs were dominantly among different physicians' specialties (emergency medicine [EM], public health professionals, surgical residents or practicing surgeons, urology residents or practicing urologists, different residencies/specialties) [77, 80, 83, 84, 86, 87, 89, 96, 99, 100, 119], with 1 study investigating nurses [107] and 1 investigating pharmacists [34].

Soares et al's [119] study compared EM trainees' and faculties' perceptions of unprofessional SM behaviors to those of state medical board directors from a prior published study [120]. They found that themes involving patient information, inappropriate communication, and discriminatory speech elicited similar probabilities of anticipated investigation by both EM and state medical board directors, compared to published data. However, compared with state medical board directors, EM physicians were less likely to anticipate that themes involving alcohol and disrespectful speech would be investigated. A study to assess changes in unprofessional content on urologists' SM was done by Koo et al [83]. Comparing the cohort in practice versus the cohort at the completion of residency, there were no significant differences in how many urologists had public Facebook accounts (70% vs 71%) or whose accounts had concerning content (43% vs 40%). Examples of concerning content included images and references to intoxication, explicit profanity, and offensive comments about patients. The presence of unprofessional content at the completion of residency strongly predicted having unprofessional content later in practice. A similar comparison was made among surgical residents and practicing surgeons [86,87]. In a study among surgical residents, 14.1% had potentially unprofessional content, and 12.2% had clearly unprofessional content. Binge drinking, sexually suggestive photos, and HIPAA violations were the most commonly found variables in the clearly unprofessional group [86]. Among attending surgeons, 10.3% had potentially unprofessional content, and 5.1% had clearly unprofessional content. Inappropriate language and sexually suggestive material were the most commonly found variables in the clearly unprofessional group [87]. Loo et al's [99] study among faculty and residents in Singapore suggested that doctors within the same residency do not necessarily have a uniform set of professional priorities regarding professionalism on SM. Data from Kesselheim et al's [80] study on pediatric residents clearly demonstrates "cognitive dissonance" in residents' approach to lapses in professionalism while using SM. More than half of the responding residents rated posting of online comments about the workplace as "completely inappropriate," yet a similar proportion estimate that residents engage in this behavior at least monthly [80]. Among nurses in China, 7.6% reported that they had "sometimes" posted identifiable patient information on SM. When asked about colleagues' online professionalism, half (50.3%) of the participants indicated that they had "sometimes" witnessed their colleagues' inappropriate SM posts and 49.5% reported "never" [107].

Among pharmacists, examples of perceived unprofessional behaviors included revealing details of personal life and activities; open complaints about the pharmacy sector, coworkers, physicians, and patients; inappropriate description of pharmacists' roles and activities; and breaches of patient confidentiality [34].

Legal Issues and Disciplinary Consequences

Unprofessional behavior on SM of HCPs can have legal consequences, potentially affecting credibility and licensure. Several studies have emphasized this issue or reflected on disciplinary legal consequences if SM are used inappropriately [61,64,78,90,111,113,115,116].

Fuoco and Leveridge [78] raised the controversy of whether medical regulatory bodies should monitor the SM activities of HCPs. In all, 94.6% of respondents agreed that physicians need to exercise caution in personal SM posting, although 57% felt that medical regulatory bodies should "stay out of [their] personal SM activities," especially those in practice for less than 10 years. Most urologists agreed that care should be taken in posting on SM sites, as unprofessional posts can put one at risk of discipline, so medicolegal guidance would be beneficial

in this aspect as well. Duke et al [115] emphasized that use of SM platforms, while potentially beneficial, can have professional and legal implications if not used appropriately in both personal and academic use. Faculty and students need to be aware that this could negatively impact their professional image and the nursing profession [115].

In Great Britain since 2013, all General Dental Council (GDC) registrants' online activities have been regulated by the GDC's SM guidelines. Failure to comply with these guidelines results in a Fitness to Practice (FtP) complaint being investigated. Documentary analysis of FtP cases from September 2013 to June 2016 revealed that 6 complaints in relation to SM were investigated. A total of 2.4% of FtP cases published on the GDC website during that period were related to breaches of the SM guidelines. All of the cases investigated were proven and upheld. Most of those named in the complaints were dental nurses, and the most common type of complaint was inappropriate Facebook comments [90]. Staud and Kearny's [113] study identified how online SM behaviors influence the licensure and enforcement practices of dental professionals. Dental boards are aware of potential online unprofessional behaviors and have implemented various consequences. Dental boards should consider developing policies to address potential online unprofessional behavior to protect the public that they serve [113]. In a recent study among Greek dental students, 75.3% of responders admitted not being aware whether the behavior of dentists on SM could result in legal sanctions [64].

Garg et al [116] conducted a survey of individual and institutional risks associated with the use of SM among residents and faculty in EM. EM residents and faculty members cause and encounter high-risk-to-professionalism events frequently while using SM; these events present significant risks to the individuals responsible and their associated institution. Some of the observations and occurrences documented in that study fall within the scope of HIPAA and put individuals and institutions at legal risk. The authors emphasize that, in addition to federal ramifications for medical institutions in regard to unprofessional conduct on SM by employees, the individuals responsible for the high-risk-to-professionalism events face state licensing consequences.

Evaluation of Existing Approaches' Effects on Promoting E-Professionalism

A total of 10 studies have tried to assess effectiveness of educational sessions or workshops incorporated in students' or residents' curriculum [46,48,52,54,55,60,63,81,89,103], and 1 study assessed the effects of formal SM instruction and policy on residents' ability to navigate case-based scenarios about online behavior in the context of professional medicine [88]. In 4 studies that included medical students as participants [46,48,60,63], educational interventions were positively accepted by students and showed a positive impact on the way they view themselves and their use of SM.

Flickinger et al [46] stated that medical educators have an opportunity not only to provide valuable guidance to students in using SM wisely but also to promote the development of professional identities by implementing SM interventions into the medical curricula [46]. In a cohort study by Gomes et al [48], 94% of medical students reported some increase in awareness, and 64% made changes to their SM behavior due to the session, reflecting the longer-term impact. Walton et al [60] preformed an exploratory pre-post study to examine the internet presence of a Canadian medical school graduating class by scanning students' public profiles on Facebook. They incorporated this information into an educational activity (3-hour long session) addressing professionalism and SM, and evaluated the impact of this activity on students' SM behavior. Repeated searches for all class members 1 month following the educational intervention revealed that many students had changed their privacy settings to further restrict public access to information on their Facebook accounts. Fewer overall students could be found by any search strategy, and in particular, there was a significant decrease in the proportion that could be found using only a simple name search. Significantly, fewer students displayed personal information or friends lists. Finally, there was a significant reduction in the number of students who openly displayed large numbers of personal photographs [60].

A similar positive effect of educational intervention was described among nursing students by Marnocha et al [52]. The study assessed effects of a peer-facilitated SM education session on changes in attitudes and knowledge among recently admitted prelicensure nursing students. Participants described plans to use a more reflective, cautious, and accountable use of SM after the intervention. Uncertain or unprofessional attitudes and knowledge showed significant improvements after the intervention [52].

One study did not find a positive correlation between educational interventions and the impact on e-professionalism among students [55]. Medical students received two 45-minute educational sessions on digital professionalism. Findings of this study suggest that isolated sessions on professionalism are not sufficient to sustain perceptions and behaviors of professionalism. Their results reflect an erosion of professionalism related to information security that occurred despite medical school and hospital-based teaching sessions to promote digital professionalism. According to Mostaghimi et al's [55] study, true alteration of trainee behavior will require a cultural shift that includes continual education; better role models; and frequent reminders for faculty, house staff, students, and staff.

A study conducted among pharmacy students showed that they are active users of peer-mediated SM learning groups. Pharmacy students have reservations regarding online professionalism and doubt the place of SM in education that includes the teacher [54].

In 3 studies that assessed the effectiveness of educational sessions on residents' perception of e-professionalism [81,89,103], the positive impact was also determined. In Khandelwal et al's [81] study, a postworkshop survey revealed that the postgraduate trainees perceived significant improvement in their understanding of e-professionalism. Compared with the preworkshop phase, residents were more comfortable defining professionalism, recognizing attributes of professionalism, describing the social contract, understanding the role of the code of conduct, and applying principles of professionalism to

challenging scenarios [81]. Similar findings were presented in Mohiuddin et al's [89] study where reflective practice-based sessions regarding the impact of SM on professionalism in surgery were well favored by the residents. Participants reported having an increased awareness to protect patient privacy and use SM more professionally [89]. Robertson et al [103] described a SM training program aimed to provide medical residents with academic and practical knowledge regarding the effective use of SM. Participants' knowledge of SM policies increased as a result of the SM training. They have also increased the ability to identify potentially inappropriate media interactions and to identify appropriate responses to such interactions, and they gained an understanding of how their actions on SM affect others [103].

One study aimed to determine the effects of formal SM instruction and policy on residents' ability to navigate case-based scenarios about online behavior in the context of professional medicine [88]. Prior SM instruction or familiarity with an SM policy were associated with improved performance on case-based questions regarding online professionalism.

Barriers to Using Social Media for Health Care Professionals

Analyzing our review sample, we have recognized that some papers highlighted important aspects of barriers that influence HCPs use of SM in the context of e-professionalism. These barriers are lack of time or time constraint, lack of knowledge or technical skills, lack of previous education or supportive institutional SM policies, ignorance to existing SM policies, and problem developing and sustaining mutual trust on SM. HCPs perceived them less as risks and more as something that keeps them away from using SM, either at all, more often, or with more quality. Lack of free time or time constraint was often recognized as a barrier [39,69,74,79,91,93,96], as well as lack of knowledge or technical skills for use of SM [33,39,76,79,93,118]. The majority of these studies that recognize the lack of time or lack of knowledge as barriers have respondents on the level of practicing HCPs. By age distribution, representatives of "millennials" or "generation Z" were not included as study participants. As shown in Adilaman et al's [74] study, this demonstrates a significant gap in SM use between younger users and mid- to late-career users. This study also found that midcareer physicians (aged 45-54 years) had statistically significantly more hesitations around joining medically geared SM sites for professional purposes, compared with those aged 25-34 years [74]. In a qualitative study among physicians by Campbell et al [76], participants expressed many levels of uncertainty about their preparedness, their impact, the potential for repercussions, and the future of physicians' presence on SM. Participants described feeling unprepared when they started using SM. Many participants described concerns such as lacking knowledge about how to use certain SM platforms versus others. Several participants felt that they were "digital immigrants" [76].

A lack of previous education about SM was emphasized in several studies [33,98,102].

A lack of SM policies was also recognized as a barrier, either as a lack of models/guidelines in how to conduct themselves online in their role as physicians, which is manifested as fear of saying the wrong thing online [76] or related mainly to being unclear about whether they are supported by their employer and professional bodies [93,97]. Contrary to that finding, even if institutions have SM policies or guidelines, HCPs acknowledged reluctant behavior regarding existing SM policies [78,85,114] or ignorance to their existence [61,64,65,100]. A lack of awareness of existing institutional SM policies was also observed for physiotherapists; 41.6% were not aware whether there was one or not [85], and half of the medical students and faculty were unaware of existing institutional SM guidelines [114].

Panahi et al [36] recognized the problem of developing and sustaining mutual trust as one of the main barriers to knowledge sharing on SM platforms [36]. Physicians trust their peers on SM in a slightly different way than in face-to-face communication. The study found that the majority of participants established trust on SM mainly through previous personal interaction, authenticity and relevancy of voice, professional standing, consistency of communication, peer recommendation, and nonanonymous and moderated sites.

Discussion

Principal Findings

A scoping review method was used to capture the latest current evidence on e-professionalism of HCPs. The 88 studies included in this scoping review cover a broad spectrum of the benefits and dangers of SM on e-professionalism for HCPs, alongside barriers perceived as threats for the limitation of SM use in the context of e-professionalism and effects of existing approaches on promoting e-professionalism. This review includes multi-perspective views from various health care professions (medical, dental, nursing, pharmacy, and physiotherapy) and from various generations of HCPs (students, residents, practicing HCPs, faculty members, and PDs/deans). Overall, the quality of the studies was satisfactory. All studies were exploratory in nature, and the findings were descriptive. Medical health professionals were involved in about three-quarters of the studies. The majority of the studies were unspecific, studying use of any type of SM or SNSs. Only Facebook or "all SM/SNSs with specific reference to Facebook" was analyzed in more than one-third of the studies. Twitter [38,44,80,91,110], Instagram [101], and YouTube [37] were specifically targeted SM/SNSs in 7 studies.

Benefits of Social Media on E-Professionalism of Health Care Professionals

Benefits of SM on e-professionalism of HCPs can be seen as improvements of established networks or possibilities for collaboration through SM sites [33, 34, 36, 38, 39, 44-46, 53, 69, 72, 76, 78, 79, 92, 97, 106-108, 117]. Besides providing the opportunity for connecting with others and sharing experiences [38,39,46,53,69,72,78,106-108], SM have enabled the creation of communities for support. The benefits of SM on e-professionalism of HCPs, identified in this scoping review as professional networking and collaboration, have been

documented in previous research for physicians [121-123], nursing profession [124,125], or other HCPs [126-129].

The benefits of peer advice, learning from peers, provision of emotional support, and identifying approaches through which physicians establish interpersonal trust on SM [36,45,92] are novel insights into the domain of e-professionalism of HCPs on SM.

Professional networking and collaborations on SM enable the development and building of professional identities for health care professions [34,38,39,45,67,73,104]. Professional identity formation among medical students now entails consideration by students about whether and how they can continue to use SM as physicians. Ruan et al's [104] study tried to define the properties and development of the digital self and its interactions with the current professional identity development theory. SM introduces new features to professional identity in the digital world. The formation of digital identity, its development, and its reconciliation with other identities were features described, and educational institutions should give more importance to navigating professional identity development. According to Cruess et al [130], students may develop "identity dissonance" when components of their identity as physicians conflict with their identity as laypersons. Research regarding identity development in SM has been primarily confined to electronic professionalism through best practice guidelines. Evolving the possibilities of SM allows HCPs to reach a large audience and can act to increase their popularity among colleagues and patients [69]. SM also creates space for self-presentation and self-promotion that has already been embraced by some HCPs, enabling them to become microcelebrities [131].

Several studies have demonstrated students' use of SM for acquiring knowledge, for gaining access to information from experts with whom they otherwise would not be able to connect, or for creating communities that can then be used as a means for supportive, professional, and social learning [44,45,58,66].

A number of studies have been conducted to investigate the ways in which health care students informally use SM for educational purposes [132]. The results identified efficient communication with educators, peer collaboration, and small group learning and sharing resources as key strengths [133]. SM has been proven to be used for educational purposes at medical schools, for example, to complement university courses [134,135]. SNSs can facilitate efficient communication, interactions, and connections among health professionals in education and training, with limitations identified as technical knowledge, professionalism, and risks of data protection [10]. Students' use of SM for health education is overwhelmingly higher in the last few years, with almost the same proportion using SM often or always [69,136]. Our findings are consistent with previous research.

Dangers of Social Media on E-Professionalism of Health Care Professionals

According to some studies in this review, loosening accountability can be seen as a danger on e-professionalism from two points of view, eroding public trust by providing poor quality of information on SM [39,106,117] and damage to professional image [43, 45, 51, 56, 57, 59, 64, 66, 68, 70, 73, 102, 106, 112].

Potential damage to professional image has been depicted by students as a concern about repercussions of their posts, on career development, or on future employment since employers are checking SM profiles of candidates [43, 45, 51, 56, 57, 59, 66, 68, 70]. In addition, it has been reported that there is more awareness of online responsibilities as students progress through their program because employers can, and at times do, use SM profiles to make hiring decisions [56,137]. Three reviewed studies investigated PDs' (medical and dental) attitudes about use of SM for admission criteria [109,110,112]. Students should be concerned about the level of professionalism presented on their profiles. Information available on SM has been already used regarding admissions to medical or nursing programs, selection for residence, or employment for over 8 years [137,138]. In 2016, the Mayo Clinic announced that it will take scholarly SM activity into account when considering academic promotion [139]. With time, it is reasonable to expect that more programs, schools, or any kind of potential employers of HCPs will use this "screening SM profiles" approach more often in the admissions process.

Compromising confidentiality concerns were described in numerous studies in this review [34, 43, 47, 57, 91, 94, 97, 107, 111, 119], especially about breaches of patient privacy or possible risks of violating HIPAA online. As previous research shows, the public availability of information on patients and physicians represents a threat to privacy [140-142], with the potential for a negative impact on patient-physician relationships [143,144]. Students and residents have a "cognitive dissonance" approach to lapses in their professionalism while using SM. It is a disconnect between what they thought they *would* do versus what they thought they *should* [47,80]. This inconsistency between attitudes and actions has been observed also elsewhere [145,146].

Traditional boundaries are blurred on many levels by online interactions. Blurred boundaries between professional and personal spheres of SM use [34, 39, 47, 51, 69, 78, 79, 91, 96, 97, 106], with concerns about exposure of one's private life, presenting details of personal life, or separating private and professional profiles, have been presented in numerous studies in this review. The recommendation that health professionals maintain a separate account with a different name, a "dual citizen approach," that maintains online professional and private identities by creating separate online profiles was introduced in 2011 by Mostaghimi and Crotty [147]. Surprisingly this issue is still prevalent. Several studies in this review investigated the purpose of SM use, whether participants mixed professional and personal information and activities on SM sites (blended profiles) or adopted a separation strategy where professional information and activities were clearly separated from personal ones (dual citizenship) [34,78,79,91,96]. Recent research shows that, for some HCPs, the risk of using SM is still a concern for the exposure of one's private life [10,39,148].

Boundaries are blurred between patients and HCPs, and between students and faculty [40, 43, 47, 59, 64, 65, 68, 70, 73, 75, 78, 82, 85, 98, 100, 105, 107, 114, 117].

Although online interaction with a patient is generally not acceptable nor endorsed, a wide range of opinions have been observed concerning this issue, ranging from one-third for medical students in Brazil [59] that find this unacceptable to 92% for senior medical students in New Zealand [70]. This disproportion in range could be explained by cultural and age differences. Some studies have demonstrated generation gap differences in friending patients, with younger age being associated with more openness to be friends with a patient [82,117]. Both students and faculty are worried that connecting via SM would blur the boundaries of the teacher-student relationship, also recognized in other studies [148,149].

Chester et al's [70] study in this review addresses a deficit in data and knowledge regarding patient-targeted Googling. This study provides a comprehensive understanding of patient-targeted Googling in concert with SNS use among senior New Zealand medical students. Results of this study show that 16.7% of respondents had conducted patient-targeted Googling. There is some evidence of an association between SNS use and likelihood of patient-targeted Googling, with high SNS users more likely to conduct patient-targeted Googling, but as the authors acknowledge, their observations were made on a small number of observations. Previous research in the United States showed that 2.3% of medical students had visited a patient's profile on an online social network [150].

Various concerns about potential professionalism implications [151-153] exist, particularly related to breaches of patient confidentiality, professional boundaries, and depiction of unprofessional behaviors. Chretien and Tuck's [14] review of online professionalism studies found that themes involving patient identifying images, inappropriate communications, and discriminatory language were consistently regarded as most inappropriate, whereas derogatory speech, images of alcohol, and partial nudity were considered only moderate to least inappropriate. Numerous studies in this review have tried to assess the extent of unprofessional behavior, posted by HCPs themselves or seen to be posted by their peers. Surveys that captured students' self-report of posted unprofessional behavior (eg, evidence of being intoxicated, illegal drug use, posting patient information, sharing clinical images inappropriately, and depictions of an illegal act) reported witnessing the investigated examples with varying frequencies [32,42,43,55,59,62,64,73,114,118].

Age difference in the term "older and wiser," meaning more cautious about posting unprofessional behavior online, was proven in studies comparing students' and faculties' online behavior [114,118]. A similar comparison was made among surgical residents and practicing surgeons with a decreasing percentage of unprofessional content among attending surgeons [86,87]. An interesting paradoxical observation from Kitsis et al's [118] study is that, although students seemed more concerned than faculty about their professional images, their online behavior did not reflect this concern. Medical students reported that they considered their online presence to be unprofessional four times more often than faculty. In view of these findings, one might expect medical students to monitor their online presence regularly. Surprisingly, they rarely reported self-monitoring and at a rate similar to the faculty. This study shows that medical students' posting of unprofessional material does not decrease during medical school and that medical students self-post and notice peers' unprofessional posts more often than faculty do [118].

Other studies have determined important differences exist in perceptions of inappropriate SM behavior among various stakeholders. Medical students often regard themes of speech, alcohol, and dress as components of online ''social identity'' rather than potential unprofessional behavior [154,155]. In contrast, patients, supervisors, and regulatory groups demonstrate more conservative views. An online survey using mock Facebook profiles found that, compared to university students, faculty and members of the public rated images significantly less appropriate [156]. Survey results showed that among students there is little consensus on what constitutes unprofessional behavior beyond the US HIPAA violations, and students have felt that posting inappropriate material on personal SM sites was "unavoidable" [156].

It seems that consensus about what constitutes unprofessional behavior, even evoked as a question since Chretien et al's [155] study in 2010, has still not been reached. There are numerous studies with examples of definitions of unprofessional behavior on SM [42,110,157,158]. Although there is no uniform consensus on what constitutes unprofessional behavior, studies most frequently associated it with online content pertaining to alcohol intoxication; substance or illegal drug use, nudity, and sexuality; demeaning content about patients, peers, educators, clinical sites, or the profession as a whole; discriminatory content; profanity; and aggressive/bullying content toward coworkers. Karveleas et al's [64] study among dental students showed that students' perceptions of and attitudes toward e-professionalism is complicated and contradictory. In their study, posting holiday pictures or wearing swimwear was categorized as unprofessional. Are these depictions of behaviors and situations unprofessional? What constitutes "potentially unprofessional behavior" has made quite a stir recently in medical scientific circles and the medical population in general.

In December 2019, a paper by Hardouin et al [159] was published investigating open, publicly available Facebook profiles of young vascular surgeons for unprofessional posts (text, images, or video content). The paper used a coding matrix, previously developed and used in other studies, for content analysis [83,84,87]. There were two distinct categories depicting e-professionalism of found content: "clearly unprofessional" and "potentially unprofessional." Three male researchers created new anonymous Facebook profiles and screened through the available data. In the "potentially unprofessional" category, images of trainees in swimwear (bikinis) screened in the research were included. This sparked controversy primarily on Twitter but also on other SM sites and mainstream media about the objectivity and bias of the researchers, reviewers, and editors, creating a hashtag #medbikini [160]. A substantial number of HCPs participated in the outraged reaction to branding posting such images or videos in bikinis as a possible sign of unprofessional behavior. They posted this content with #medbikini and their disapproval of such a label and referred to the gender bias of the researchers, questioning possibly outdated norms of behavior for HCPs [161]. This ultimately led

to the retraction of the paper and publication of a "Retraction notice" by the editors of the Journal of Vascular Surgery [162].

In a recently published paper by Pronk et al [163] that studied all levels of medical professionals (students, residents, and specialists), the authors found that all investigated groups perceived information or pictures to be unprofessional related to alcohol abuse, partying, and sexually suggestive posts, creating a dissonance between the #medbiniki movement's perception of professionalism and collected data [163]. However, they argue that some of the participants' opinions could have changed due to the debate initiated by the #medbikini movement, which occurred after their data collection.

Another recent study by Meira et al [164] investigating professionalism perception of orthodontist through exposure of laypeople, dental students, and dentists to images usually found on Instagram found that images related to social and family relationships were associated with lower scores regarding the perception of professional credibility for all groups [164]. They argued that their results indicate that personal images, possibly because they are not related to the professional context, contribute little toward the professional image of orthodontists on Instagram.

Unprofessional behavior on SM of HCPs can have legal consequences, potentially affecting credibility and licensure. Several studies have emphasized this issue or reflected on possible professional consequences if SM are used inappropriately [61,64,78,90,111,113,115,116].

Previous research has described associations of specific SM behaviors with the risk of investigation and subsequent disciplinary action by regulatory agencies by state medical boards and reported that online violations of professionalism by physicians were quite common and often led to disciplinary actions [120,165]. The consequences in the breach of privacy in the nursing profession can be severe and may lead to civil or criminal penalties [166]. Recent studies have also recognized that consequences of unprofessional online SM use can result in expulsion, lawsuits, job loss, and permanently damaged professional reputations [167]. This can also result in inaction or lack of use of SM for beneficial purposes, as the fear of legal issues can hinder use. This was recognized in a recent study by Al-Khalifa et al [168] where on a population of Saudi Arabian dentist only 41% of them were inclined to give online consultations. They argued the rest were possibly fearing potential legal ramifications. In an age of social distancing due to COVID-19, this could lead to patients not receiving information or care that they need and could have possibly gotten through online contact.

Recommendations for E-Professionalism Curriculum Changes

Ten years ago, many schools lacked policies specific to SM use [151], but nowadays schools have developed specific guidelines [169,170]. Guidelines are also available from numerous professional societies [171-178], and a recent review about available guidelines from nine medical international bodies has been published [179]. Previous work on health care education interventions and experiences has noted how learners may be motivated to reduce the hazards of SM, revise SM confidentiality settings, or even terminate SM involvement upon realizing that online postings may have an enduring presence [180,181].

Effectiveness of educational interventions about e-professionalism or impact of existing SM policies has been recognized in this review, since numerous studies explored educational interventions for promoting e-professionalism [46,48,52,54,55,60,63,81,89,103].

On an educational level for students, recommendations are to include a variety of e-professionalism topics into a curriculum to provide students with a clear picture of what constitutes professional violations on SM and assist them in distinguishing between personal and professional personas online [42,43,47,49,50,53,54,64,103,111,114]. Hsieh et al's [63] study demonstrates the possibility of how SM can be used as a learning platform for professionalism, enabling students a virtual space in which to share positive examples that reflect the authentic experience in a clinical environment. Our previous findings demonstrate that the perception of unprofessional behavior varies among HCPs, mostly associated with age of the participants [86,109,114,118]. Similar findings were confirmed also for health science students who struggle with the concepts associated with professionalism [182]. Teaching professionalism in general offers challenges for educators, and these challenges are amplified when the topic moves into cyberspace, where students are digital natives and faculty are generally digital immigrants [136]. Several studies in this review have recognized the need to include students in the development of guidelines [47] or to assist in education with somebody of their age group, providing personal experiences and more of a "nonauthoritative" approach [49].

O'Sullivan et al [32] have also recognized the importance of schools using an evidence-based approach to policy creation and to involve students in the process of the creation of these policies. A recent study by Wissinger and Stiegler [183] also highlights the importance of formal integration of e-professionalism into the health care curricula to prepare students for situations they will face once employed. By placing the responsibility of learning e-professionalism inside the walls of academia, students are prepared to take control of their online identity and craft a persona that represents their professional image [167]. As Chretien and Kind [183] described it, a victory for online professionalism would be providing trainees with tools and guidance needed to ascend on the SM hierarchy pyramid of needs, from public trust to discovery.

Similar recommendations were described in this review for residents, with important issues that must be addressed during curriculum development: integrate trainees as educators, encourage peer-to-peer regulation, and provide opportunity for reflection. Effective educational interventions for teaching online professionalism must include the skills necessary for residents not only to recognize inappropriate behavior on SM but also to learn how to address it themselves [80,103]. There is a qualitative distinction between disseminating guidelines and formally integrating SM instruction into medical curricula, which should become imperative for HCP education,

undergraduate or graduate level, or continuing medical education [42]. Similar conclusions were made in a systematic review of SM in residency [184]. Economides et al's [184] review depicts evolving perceptions and a paradigm shift, where a growing body of literature is now focusing on promoting responsible SM use, examining how SM training can enhance professional growth and academic scholarship. As the tone of the dialogue shifts from trepidation to interest or even to enthusiasm, it is clear that there is a need for formalized standards and education on SM use established within the trainee's curriculum.

Barriers to Social Media Use for Health Care Professionals

Analyzing our review sample, we have recognized that some papers highlighted important aspects of barriers that influence HCPs use of SM in context of e-professionalism. A lack of free time or time constraint was often recognized as a barrier [39,69,74,79,91,93,96], as well as a lack of knowledge or technical skills for use of SM [33,39,76,79,93,118]. Studies in this review with "lack of time" or "lack of knowledge" barriers had respondents that were practicing HCPs and older HCPs; representatives of "millennials" or "generation Z" were not included as study participants. This demonstrates a significant gap in SM use between younger users and mid- to late-career users [74,82]. This is consistent with previous research demonstrating that in addition to the practical barriers to adoption of SM in the professional realm, a generation gap exists, with millennials using SM for contact and information far more frequently than members of generation X and baby boomers [150]. Similar results can be found in Chan et al's [10] systematic review where identified positive predictors of use of SNSs for professional purposes were younger age (20-39 years), fewer years of professional experience (0-10 years), and lower rank, such as residents.

Our results show that even though a lack of SM policies was recognized as a barrier, even if institutions have SM policies or guidelines, HCPs acknowledged reluctant behavior regarding existing SM policies [78,85,114] or ignorance to their existence [61,64,65,100]. This should be considered as a warning to increase awareness on this matter, as SM will continue to be increasingly ubiquitous and integrated in health care. As Parsi and Elster [185] note, "if we fail to engage this technology constructively, we will lose an important opportunity to expand the application of medical professionalism within contemporary society." Since the SM world is changing so fast, adopting novel approaches to existing SM policies becomes essential. As Kerr et al [101] suggest, it is imperative for nursing education, professional regulatory bodies, and employers to develop more robust and dynamic policies and guidelines related to the appropriate use of SM within the profession, especially with the growing presence of web-based HCP microcelebrities [131].

Comparison With Prior Work

Compared to other literature reviews published on related topics, this scoping review is the first to capture original research about e-professionalism in terms of methods, subjects, and themes since Chretien and Tuck [14], who conducted a synthetic review to characterize the original peer-reviewed research on online professionalism of medical students, residents, or physicians. The review included 32 studies and recognized general areas of online professionalism (use and privacy, assessment of unprofessional behavior, consensus-gathering of what constitutes unprofessional or inappropriate behaviors, and education and policies) with no clear separation between challenges or benefits and addressed only online professionalism of medical students, residents, or physicians. Other reviews presented a full spectrum of SM-related challenges and opportunities in the context of medical professionalism of diverse types of HCPs [15,16] or in the context of SM as an emerging tool in education [132,186], but these studies were conducted several years ago.

Similar conclusions were made in other research. Although there exist concerns about misuse of SM and violation of e-professionalism by HCPs, SM can also be used constructively as a tool for professional development; as a means of accessing information, marketing practices and services, job opportunities; and as a means of sharing or adding opinions on issues of interest to HCPs and to other like-minded individuals online [44,187].

Ventola [121] recognized benefits, risks, and best practices for HCPs. He concluded that SM can provide considerable benefits in professional networking or collaboration, professional education, patient's care, education, and health programs. All these benefits of SM were recognized in this scoping review as well. According to Ventola [121], there are some risks related to poor quality of information, damage to professional image, breaches of patient privacy, violation of the patient-HCP boundary, and licensing and legal issues, which was also recognized in this scoping review. Likewise, risks such as privacy and accuracy of information, compromising confidentiality, eroding public trust, and loosening accountability were presented in previous research [156,188,189].

Gaps in the Literature and Potential Areas for Further Research

This review demonstrates dominance of Facebook in research done so far. With the rapid evolution of SM, future insights should be more oriented toward new and emerging SM sites. Instagram has gained an enormous following with new features like "Stories" and "Reels" within the SM itself, which are completely scientifically unexplored. Snapchat and TikTok have also gained a substantial following, especially among the student population. TikTok did not even exist in 2014; nowadays, TikTok has 689 million users worldwide [20]. They function on a completely different set of parameters, being that the content is time limited. New research should consider how to approach the youngest generation of HCPs who are using these SM sites and how to design a novel study methodology to gain insight, due to the time limitation of the content.

Gaining popularity on SM is not only reserved for adolescents and young adults. Creation of influencers who are HCPs can affect perception of the professional image, either positively or negatively. This has been rarely analyzed so far.

Geographical locations may affect the generalization of findings in research on SM use. Asian countries have regionally oriented

SM and SNSs like WeChat not used in European countries or the United States. Cultural differences should also be considered.

The curriculum implementation of SM guidelines and educational efforts are also there to be evaluated. As the diversity of such actions is apparent, efficiency is key in getting the proper message to a generation of "millennials" with a short attention span.

The COVID-19 pandemic has caused much of the world's population to isolate itself and many of us to shift our lives to digital tech platforms, especially SM and SNSs, all experiencing strong growth. Previous research has shown that more people are relying on SM to find and share health information during times of crisis. Future research should investigate how the pandemic affected our e-professional behavior. We are experiencing an unprecedented time in health care and education due to the COVID-19 pandemic, so the use of SM in patient/HCP communication and student education should also be explored in more detail.

SM can also be used in marketing and self-promotion [190]. Dental medicine is much more open to such actions, with medical professionals showing a lack of interest and being more worried about legal ramifications. Research into the reasons of such divergence and insight into the negative attitudes is essential for creating a platform for implementing SM in a positive and professional manner.

Findings of this research confirm the dominance of medical students or physicians as a study population of HCPs in the context of e-professionalism. Future research could be done to further investigate other types of HCPs with an emphasis on the specifics of each profession regarding their SM potential and use. Comparison among different types of HCPs would add novel insights to the field of e-professionalism.

Limitations of the Review

We acknowledge that scoping reviews have several limitations, but a scoping review allowed us to gain a wide-ranging understanding of the impact of SM on e-professionalism of HCPs. Research into SM is rapidly growing, and this scoping review is a snapshot of the latest current evidence on e-professionalism of HCPs.

There might be a selection bias (failure to search in additional potentially relevant databases to which the university does not have access) and a publication bias (we only searched in 3 databases, we did not extensively search for gray literature, and our search was limited to the English language). All studies were exploratory in nature, and the findings were descriptive. The questionnaires adopted in the surveys were mostly developed by the researchers, where validation mostly was not done. Research on SM is growing so fast that evidence may have been published in electronic media or platforms not indexed through the academic databases. Thus, findings in this review are limited to research published in traditional peer-reviewed journals only.

Conclusions

A scoping review was conducted that included 88 studies, offering current evidence on e-professionalism of HCPs. Almost all studies were found to be of adequate quality. Findings in reviewed studies indicate the existence of both benefits of SM on e-professionalism such as professional networking and collaboration, training, and education, and, on the other hand, the dangers of SM, such as loosening of accountability, compromising doctor-patient confidentiality, blurred professional boundaries, depiction of unprofessional behavior on SM, and legal consequences.

Even though there are some barriers recognized, this review has highlighted existing recommendations for including e-professionalism in educational curricula of HCPs. Based on all evidence provided, this review provided new insights and guides for future research on this area. There is a clear need for robust research to investigate new emerging SM platforms, the efficiency of guidelines and educational interventions, and the specifics of each profession regarding their SM potential and use.

Acknowledgments

The authors thank Prof Howard H Goldman and Alejandro Ortiz, PhD, School of Medicine, University of Maryland for their support. This study was funded by the Croatian Science Foundation under project UIP-05-2017 "Dangers and benefits of social networks: E-Professionalism of healthcare professionals – SMePROF."

Authors' Contributions

TVR, JV, and DJ conceived and designed the review, and TVR coordinated it. TVR, JV, and LMP were involved in developing the search strategy, and TVR, JV, LMP, DJ, KS, and MM extracted the data. Data analyses were undertaken by TVR and JV, whereas data interpretation was done by TVR, JV, LMP, and MM. TVR drafted the review. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.

[PDF File (Adobe PDF File), 210 KB - jmir_v23i11e25770_app1.pdf]

Multimedia Appendix 2
Summary of the reviewed studies (n=88).
[PDF File (Adobe PDF File), 479 KB - jmir_v23i11e25770_app2.pdf]

Multimedia Appendix 3
List of the excluded studies and the reasons for exclusion.
[DOCX File , 102 KB - jmir_v23i11e25770_app3.docx]

Multimedia Appendix 4
Quality of the reviewed studies.
[DOCX File , 55 KB - jmir_v23i11e25770_app4.docx]

References

1. Kemp S. Digital 2019: global internet use accelerates. We Are Social. 2019 Jan 30. URL: https://wearesocial.com/uk/blog/2019/01/digital-in-2019-global-internet-use-accelerates/ [accessed 2020-10-22]
2. Social media fact sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-10-22]
3. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Klassen TP, et al. Social media use by health care professionals and trainees: a scoping review. Acad Med 2013 Sep;88(9):1376-1383. [doi: 10.1097/ACM.0b013e31829eb91c] [Medline: 23887004]
4. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
5. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
6. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
7. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
8. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
9. Coiera E. Social networks, social media, and social diseases. BMJ 2013 May 22;346:f3007. [doi: 10.1136/bmj.f3007] [Medline: 23697672]
10. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):e117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
11. Ashton KS. Teaching nursing students about terminating professional relationships, boundaries, and social media. Nurse Educ Today 2016 Feb;37:170-172. [doi: 10.1016/j.nedt.2015.11.007] [Medline: 26602238]
12. Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age. Curr Pharm Teaching Learning 2009 Dec;1(2):66-70. [doi: 10.1016/j.cptl.2009.10.001]
13. Greysen SR, Kind T, Chretien KC. Online professionalism and the mirror of social media. J Gen Intern Med 2010 Nov;25(11):1227-1229 [FREE Full text] [doi: 10.1007/s11606-010-1447-1] [Medline: 20632121]
14. Chretien KC, Tuck MG. Online professionalism: a synthetic review. Int Rev Psychiatry 2015 Apr;27(2):106-117. [doi: 10.3109/09540261.2015.1004305] [Medline: 25804627]
15. Gholami-Kordkheili F, Wild V, Strech D. The impact of social media on medical professionalism: a systematic qualitative review of challenges and opportunities. J Med Internet Res 2013 Aug 28;15(8):e184 [FREE Full text] [doi: 10.2196/jmir.2708] [Medline: 23985172]
16. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
17. Kemp S. Digital 2014: global digital statshot December 2014. DataReportal. URL: https://datareportal.com/reports/digital-2014-global-digital-statshot-december-2014 [accessed 2021-04-19]
18. Kemp S. Digital 2014: global social media users pass 2 billion mark. DataReportal. URL: https://datareportal.com/reports/digital-2014-global-social-media-users-pass-2-billion-mark [accessed 2021-04-19]

19. Kemp S. Digital 2020: October global statshot. DataReportal. URL: https://datareportal.com/reports/digital-2020-october-global-statshot [accessed 2021-04-19]
20. Tankovska H. Number of social network users worldwide from 2017 to 2025. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2021-04-19]
21. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
22. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol 2018 Jan 10;18(1):5 [FREE Full text] [doi: 10.1186/s12874-017-0468-4] [Medline: 29316881]
23. Tricco AC, Soobiah C, Antony J, Cogo E, MacDonald H, Lillie E, et al. A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. J Clin Epidemiol 2016 May;73:19-28. [doi: 10.1016/j.jclinepi.2015.08.030] [Medline: 26891949]
24. Peters M, Godfrey C, McInerney P, Munn Z, Trico A, Khalil H. Chapter 11: scoping reviews. JBI Global Wiki. URL: https://jbi-global-wiki.refined.site/space/MANUAL/3283910770/Chapter+11%3A+Scoping+reviews [accessed 2021-04-09]
25. Peters MDJ, 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]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
27. Vukusic Rukavina T, Viskic J, Machala Poplasen L, Relic D, Marelic M, Jokic D, et al. Dangers and benefits of social media on e-professionalism of healthcare professionals: scoping review. OSF. 2021 Sep 23. URL: https://osf.io/yr8tw [accessed 2021-11-10]
28. Neuendorf K. The Content Analysis Guidebook 2nd edition. London: SAGE Publications; 2017.
29. Freelon DG. ReCal: intercoder reliability calculation as a web service. Int J Internet Sci 2010;5(1):33 [FREE Full text]
30. Quality assessment tool for observational cohort and cross-sectional studies. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort [accessed 2019-04-27]
31. CASP checklists. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2019-04-11]
32. O'Sullivan E, Cutts E, Kavikondala S, Salcedo A, D'Souza K, Hernandez-Torre M, et al. Social media in health science education: an international survey. JMIR Med Educ 2017 Jan 04;3(1):e1 [FREE Full text] [doi: 10.2196/mededu.6304] [Medline: 28052842]
33. D'Souza K, Henningham L, Zou R, Huang J, O'Sullivan E, Last J, et al. Attitudes of health professional educators toward the use of social media as a teaching tool: global cross-sectional study. JMIR Med Educ 2017 Aug 04;3(2):e13 [FREE Full text] [doi: 10.2196/mededu.6429] [Medline: 28778841]
34. Benetoli A, Chen TF, Schaefer M, Chaar B, Aslani P. Pharmacists' perceptions of professionalism on social networking sites. Res Social Adm Pharm 2017;13(3):575-588. [doi: 10.1016/j.sapharm.2016.05.044] [Medline: 27423783]
35. Levati S. Professional conduct among registered nurses in the use of online social networking sites. J Adv Nurs 2014 Oct;70(10):2284-2292. [doi: 10.1111/jan.12377] [Medline: 24617801]
36. Panahi S, Watson J, Partridge H. Fostering interpersonal trust on social media: physicians' perspectives and experiences. Postgrad Med J 2016 Feb;92(1084):70-73. [doi: 10.1136/postgradmedj-2015-133270] [Medline: 26627975]
37. Ahmad T, Sattar K, Akram A. Medical professionalism videos on YouTube: content exploration and appraisal of user engagement. Saudi J Biol Sci 2020 Sep;27(9):2287-2292 [FREE Full text] [doi: 10.1016/j.sjbs.2020.06.007] [Medline: 32884409]
38. Borgmann H, DeWitt S, Tsaur I, Haferkamp A, Loeb S. Novel survey disseminated through Twitter supports its utility for networking, disseminating research, advocacy, clinical practice and other professional goals. Can Urol Assoc J 2015;9(9-10):E713-E717 [FREE Full text] [doi: 10.5489/cuaj.3014] [Medline: 26664662]
39. Nikiphorou E, Studenic P, Ammitzbøll CG, Canavan M, Jani M, Ospelt C, EMEUNET. Social media use among young rheumatologists and basic scientists: results of an international survey by the Emerging EULAR Network (EMEUNET). Ann Rheum Dis 2017 Apr;76(4):712-715. [doi: 10.1136/annrheumdis-2016-209718] [Medline: 27797750]
40. Alkhateeb F, Alameddine S, Attarabeen O, Latif D, Osolin S, Khanfar N, et al. Pharmacy students' use of social media sites and perception toward Facebook use. Arch Pharm Pract 2015;6(4):77. [doi: 10.4103/2045-080x.165134]
41. Bagley JE, DiGiacinto D, Lawyer J, Anderson MP. Health Care Students Who Frequently Use Facebook Are Unaware of the Risks for Violating HIPAA Standards. J Diagn Med Sonography 2014 Apr 09;30(3):114-120. [doi: 10.1177/8756479314530509]
42. Barlow CJ, Morrison S, Stephens HO, Jenkins E, Bailey MJ, Pilcher D. Unprofessional behaviour on social media by medical students. Med J Aust 2015 Dec 14;203(11):439. [doi: 10.5694/mja15.00272] [Medline: 26654611]
43. Barnable A, Cunning G, Parcon M. Nursing students' perceptions of confidentiality, accountability, and e-professionalism in relation to Facebook. Nurse Educ 2018;43(1):28-31. [doi: 10.1097/NNE.0000000000000441] [Medline: 28857955]

44. Chretien KC, Tuck MG, Simon M, Singh LO, Kind T. A digital ethnography of medical students who use Twitter for professional development. J Gen Intern Med 2015 Nov;30(11):1673-1680 [FREE Full text] [doi: 10.1007/s11606-015-3345-z] [Medline: 25952652]
45. Ferguson C, DiGiacomo M, Saliba B, Green J, Moorley C, Wyllie A, et al. First year nursing students' experiences of social media during the transition to university: a focus group study. Contemp Nurse 2016 Oct;52(5):625-635. [doi: 10.1080/10376178.2016.1205458] [Medline: 27334672]
46. 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]
47. George DR, Navarro AM, Stazyk KK, Clark MA, Green MJ. Ethical quandaries and Facebook use: how do medical students think they (and their peers) should (and would) act? AJOB Empirical Bioethics 2014 Apr 14;5(2):68-79. [doi: 10.1080/23294515.2013.864344]
48. Gomes AW, Butera G, Chretien KC, Kind T. The development and impact of a social media and professionalism course for medical students. Teach Learn Med 2017;29(3):296-303. [doi: 10.1080/10401334.2016.1275971] [Medline: 28272900]
49. 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]
50. Jawaid M, Khan MH, Bhutto SN. Pak J Med Sci 2015;31(1):209-213 [FREE Full text] [doi: 10.12669/pjms.311.5643] [Medline: 25878645]
51. Kenny P, Johnson IG. Social media use, attitudes, behaviours and perceptions of online professionalism amongst dental students. Br Dent J 2016 Nov 18;221(10):651-655. [doi: 10.1038/sj.bdj.2016.864] [Medline: 27857111]
52. Marnocha S, Marnocha M, Cleveland R, Lambie C, Limberg CY, Wnuk J. A peer-delivered educational intervention to improve nursing student cyberprofessionalism. Nurse Educ 2017;42(5):245-249. [doi: 10.1097/NNE.0000000000000368] [Medline: 28252547]
53. Mather C, Cummings E, Nichols L. Social media training for professional identity development in undergraduate nurses. Stud Health Technol Inform 2016;225:344-348. [Medline: 27332219]
54. Mawdsley A. Pharmacy students' perceptions of social media in education. Pharm Education 2015;15(1):108-110 [FREE Full text]
55. Mostaghimi A, Olszewski AE, Bell SK, Roberts DH, Crotty BH. Erosion of digital professionalism during medical students' core clinical clerkships. JMIR Med Educ 2017 May 03;3(1):e9 [FREE Full text] [doi: 10.2196/mededu.6879] [Medline: 28468745]
56. Nason KN, Byrne H, Nason GJ, O'Connell B. An assessment of professionalism on students' Facebook profiles. Eur J Dent Educ 2018 Feb;22(1):30-33 [FREE Full text] [doi: 10.1111/eje.12240] [Medline: 27735108]
57. Ness GL, Sheehan AH, Snyder ME. Graduating student pharmacists' perspectives on e-professionalism and social media: qualitative findings. J Am Pharm Assoc (2003) 2014;54(2):138-143. [doi: 10.1331/JAPhA.2014.13188] [Medline: 24632929]
58. Nicolai L, Schmidbauer M, Gradel M, Ferch S, Antón S, Hoppe B, et al. Facebook groups as a powerful and dynamic tool in medical education: mixed-method study. J Med Internet Res 2017 Dec 22;19(12):e408 [FREE Full text] [doi: 10.2196/jmir.7990] [Medline: 29273572]
59. Rocha PN, de Castro NAA. Opinions of students from a Brazilian medical school regarding online professionalism. J Gen Intern Med 2014 May;29(5):758-764 [FREE Full text] [doi: 10.1007/s11606-013-2748-y] [Medline: 24395103]
60. Walton JM, White J, Ross S. What's on YOUR Facebook profile? Evaluation of an educational intervention to promote appropriate use of privacy settings by medical students on social networking sites. Med Educ Online 2015;20:28708 [FREE Full text] [doi: 10.3402/meo.v20.28708] [Medline: 26198434]
61. Dobson E, Patel P, Neville P. Perceptions of e-professionalism among dental students: a UK dental school study. Br Dent J 2019 Jan 11;226(1):73-78. [doi: 10.1038/sj.bdj.2019.11] [Medline: 30631197]
62. Gupta S, Singh S, Dhaliwal U. Visible Facebook profiles and e-professionalism in undergraduate medical students in India. J Educ Eval Health Prof 2015;12:50. [doi: 10.3352/jeehp.2015.12.50] [Medline: 26582630]
63. Hsieh J, Kuo L, Wang Y. Learning medical professionalism - the application of appreciative inquiry and social media. Med Educ Online 2019 Dec;24(1):1586507 [FREE Full text] [doi: 10.1080/10872981.2019.1586507] [Medline: 30831060]
64. Karveleas I, Kyriakouli A, Koukou M, Koufatzidou M, Kalogirou E, Tosios KI. The relationship between Facebook behaviour and e-professionalism: a questionnaire-based cross-sectional study among Greek dental students. Eur J Dent Educ 2021 Feb;25(1):151-158. [doi: 10.1111/eje.12585] [Medline: 32780448]
65. Knott PN, Wassif HS. Older and wiser? First year BDS graduate entry students and their views on using social media and professional practice. Br Dent J 2018 Sep 14;225(5):437-440. [doi: 10.1038/sj.bdj.2018.745] [Medline: 30168814]
66. Sadd R. Student nurse attitudes and behaviours when using social network sites. Stud Health Technol Inform 2019 Aug 21;264:1342-1346. [doi: 10.3233/SHTI190445] [Medline: 31438144]
67. West CA, Wagner JM, Greenberg SB, Buck E, Hsieh P, Horn K, et al. Examining medical students' social media beliefs and behaviors and their relationship to professional identity. Med Sci Educator 2018 Mar 26;28(2):389-399. [doi: 10.1007/s40670-018-0562-1]

68. Yang Y, Jeong E, Je NK, Jee J, Yoo JC, Choi EJ. An Analysis of Pharmacy Students' Social Networking Service Activities and Perceptions Regarding E-Professionalism under the Newly Implemented 6-year Pharmacy Educational System in South Korea. Indian J Pharm Education Res 2016 Jan 01;50(1):63-69. [doi: 10.5530/ijper.50.1.9]
69. Avcı K, Çelikden SG, Eren S, Aydenizöz D. Assessment of medical students' attitudes on social media use in medicine: a cross-sectional study. BMC Med Educ 2015 Feb 15;15:18 [FREE Full text] [doi: 10.1186/s12909-015-0300-y] [Medline: 25890252]
70. Chester AN, Walthert SE, Gallagher SJ, Anderson LC, Stitely ML. Patient-targeted Googling and social media: a cross-sectional study of senior medical students. BMC Med Ethics 2017 Dec 04;18(1):70 [FREE Full text] [doi: 10.1186/s12910-017-0230-9] [Medline: 29202840]
71. Gettig JP, Noronha S, Graneto J, Obucina L, Christensen KJ, Fjortoft NF. Examining health care students' attitudes toward e-professionalism. Am J Pharm Educ 2016 Dec 25;80(10):169 [FREE Full text] [doi: 10.5688/ajpe8010169] [Medline: 28179718]
72. Hinojo-Lucena F, Aznar-Díaz I, Cáceres-Reche M, Romero-Rodríguez J. Use of social networks for international collaboration among medical students. Educación Médica 2020 Mar;21(2):137-141. [doi: 10.1016/j.edumed.2018.08.009]
73. Nyangeni T, Du Rand S, Van Rooyen D. Perceptions of nursing students regarding responsible use of social media in the Eastern Cape. Curationis 2015 Jul 24;38(2):1496 [FREE Full text] [doi: 10.4102/curationis.v38i2.1496] [Medline: 26244464]
74. Adilman R, Rajmohan Y, Brooks E, Urgoiti GR, Chung C, Hammad N, et al. Social media use among physicians and trainees: results of a national medical oncology physician survey. J Oncol Pract 2016 Jan;12(1):79-80, e52. [doi: 10.1200/JOP.2015.006429] [Medline: 26443837]
75. Alshakhs F, Alanzi T. The evolving role of social media in health-care delivery: measuring the perception of health-care professionals in Eastern Saudi Arabia. J Multidiscip Healthc 2018;11:473-479. [doi: 10.2147/JMDH.S171538] [Medline: 30275699]
76. Campbell L, Evans Y, Pumper M, Moreno MA. Social media use by physicians: a qualitative study of the new frontier of medicine. BMC Med Inform Decis Mak 2016 Jul 15;16:91 [FREE Full text] [doi: 10.1186/s12911-016-0327-y] [Medline: 27418201]
77. Dawkins R, King WD, Boateng B, Nichols M, Desselle BC. Pediatric residents' perceptions of potential professionalism violations on social media: a US national survey. JMIR Med Educ 2017 Jan 31;3(1):e2 [FREE Full text] [doi: 10.2196/mededu.5993] [Medline: 28143804]
78. Fuoco M, Leveridge MJ. Early adopters or laggards? Attitudes toward and use of social media among urologists. BJU Int 2015 Mar;115(3):491-497. [doi: 10.1111/bju.12855] [Medline: 24981237]
79. Irfan KS, Farhana I, Eiad AF, Nassr AM, Al Mohammed AQ, Maya N, et al. Family physicians' utility of social media: a survey comparison among family medicine residents and physicians. Afr Health Sci 2018 Sep;18(3):817-827 [FREE Full text] [doi: 10.4314/ahs.v18i3.41] [Medline: 30603016]
80. Kesselheim JC, Schwartz A, Belmonte F, Boland KA, Poynter S, Batra M, Association of Pediatric Program Directors Longitudinal Educational Assessment Research Network (APPD LEARN) Study Group on Social MediaProfessionalism. A national survey of pediatric residents' professionalism and social networking: implications for curriculum development. Acad Pediatr 2016 Mar;16(2):110-114. [doi: 10.1016/j.acap.2015.12.004] [Medline: 26718877]
81. Khandelwal A, Nugus P, Elkoushy MA, Cruess RL, Cruess SR, Smilovitch M, et al. How we made professionalism relevant to twenty-first century residents. Med Teach 2015;37(6):538-542. [doi: 10.3109/0142159X.2014.990878] [Medline: 25594336]
82. Klee D, Covey C, Zhong L. Social media beliefs and usage among family medicine residents and practicing family physicians. Fam Med 2015 Mar;47(3):222-226 [FREE Full text] [Medline: 25853534]
83. Koo K, Bowman MS, Ficko Z, Gormley EA. Older and wiser? Changes in unprofessional content on urologists' social media after transition from residency to practice. BJU Int 2018 Aug;122(2):337-343. [doi: 10.1111/bju.14363] [Medline: 29694713]
84. Koo K, Ficko Z, Gormley EA. Unprofessional content on Facebook accounts of US urology residency graduates. BJU Int 2017 Jun;119(6):955-960. [doi: 10.1111/bju.13846] [Medline: 28393475]
85. Laliberté M, Beaulieu-Poulin C, Campeau Larrivée A, Charbonneau M, Samson É, Ehrmann Feldman D. Current uses (and potential misuses) of Facebook: an online survey in physiotherapy. Physiother Can 2016;68(1):5-12 [FREE Full text] [doi: 10.3138/ptc.2014-41] [Medline: 27504042]
86. Langenfeld SJ, Cook G, Sudbeck C, Luers T, Schenarts PJ. An assessment of unprofessional behavior among surgical residents on Facebook: a warning of the dangers of social media. J Surg Educ 2014;71(6):e28-e32. [doi: 10.1016/j.jsurg.2014.05.013] [Medline: 24981657]
87. Langenfeld SJ, Sudbeck C, Luers T, Adamson P, Cook G, Schenarts PJ. The glass houses of attending surgeons: an assessment of unprofessional behavior on Facebook among practicing surgeons. J Surg Educ 2015;72(6):e280-e285. [doi: 10.1016/j.jsurg.2015.07.007] [Medline: 26276300]
88. Lefebvre C, Mesner J, Stopyra J, O'Neill J, Husain I, Geer C, et al. Social media in professional medicine: new resident perceptions and practices. J Med Internet Res 2016 Jun 09;18(6):e119 [FREE Full text] [doi: 10.2196/jmir.5612] [Medline: 27283846]

89. Mohiuddin Z, Shahid H, Shuaib W. Social media impact: utility of reflective approach in the practice of surgery. Indian J Surg 2015 Dec;77(Suppl 3):893-898 [FREE Full text] [doi: 10.1007/s12262-014-1056-z] [Medline: 27011477]
90. Neville P. Social media and professionalism: a retrospective content analysis of Fitness to Practise cases heard by the GDC concerning social media complaints. Br Dent J 2017 Sep 08;223(5):353-357. [doi: 10.1038/sj.bdj.2017.765] [Medline: 28883584]
91. Patel SS, Hawkins CM, Rawson JV, Hoang JK. Professional social networking in radiology: who is there and what are they doing? Acad Radiol 2017 May;24(5):574-579. [doi: 10.1016/j.acra.2016.09.026] [Medline: 28153576]
92. Pereira I, Cunningham AM, Moreau K, Sherbino J, Jalali A. Thou shalt not tweet unprofessionally: an appreciative inquiry into the professional use of social media. Postgrad Med J 2015 Oct;91(1080):561-564 [FREE Full text] [doi: 10.1136/postgradmedj-2015-133353] [Medline: 26294333]
93. Scragg B, Shaikh S, Shires G, Stein Hodgins J, Mercer C, Robinson L, et al. An exploration of mammographers' attitudes towards the use of social media for providing breast screening information to clients. Radiography (Lond) 2017 Aug;23(3):249-255. [doi: 10.1016/j.radi.2017.04.004] [Medline: 28687294]
94. Surani Z, Hirani R, Elias A, Quisenberry L, Varon J, Surani S, et al. Social media usage among health care providers. BMC Res Notes 2017 Nov 29;10(1):654 [FREE Full text] [doi: 10.1186/s13104-017-2993-y] [Medline: 29187244]
95. Wagner JP, Cochran AL, Jones C, Gusani NJ, Varghese TK, Attai DJ. Professional use of social media among surgeons: results of a multi-institutional study. J Surg Educ 2018;75(3):804-810. [doi: 10.1016/j.jsurg.2017.09.008] [Medline: 28964746]
96. Weijs C, Majowicz S, Coe JB, Desmarais S, Jones-Bitton A. The personal use of Facebook by public health professionals in Canada: implications for public health practice. J Commun Healthcare 2017 Jan 12;10(1):8-15. [doi: 10.1080/17538068.2016.1274846]
97. Justinia T, Alyami A, Al-Qahtani S, Bashanfar M, El-Khatib M, Yahya A, et al. Social media and the orthopaedic surgeon: a mixed methods study. Acta Inform Med 2019 Mar;27(1):23-28 [FREE Full text] [doi: 10.5455/aim.2019.27.23-28] [Medline: 31213739]
98. Lee YJ, Hwang J, Lee SI, Woo S, Hahn SW, Koh S. Impact of experience of psychiatrists and psychiatry residents regarding electronic communication and social networking on internet use patterns: a questionnaire survey for developing e-professionalism in South Korea. BMC Med Educ 2019 Nov 08;19(1):411 [FREE Full text] [doi: 10.1186/s12909-019-1771-z] [Medline: 31703677]
99. Loo ME, Wong B, Lee YM. Evaluating the appropriateness of Facebook posts – What do faculty and residents consider? Asia Pacific Scholar 2020 Sep 1;5(3):71-82. [doi: 10.29060/taps.2020-5-3/oa2226]
100. Low JM, Tan MY, Joseph R. Doctors and social media: knowledge gaps and unsafe practices. Singapore Med J 2020 Apr 21:1. [doi: 10.11622/smedj.2020067] [Medline: 32312027]
101. Kerr H, Booth R, Jackson K. Exploring the characteristics and behaviors of nurses who have attained microcelebrity status on Instagram: content analysis. J Med Internet Res 2020 May 26;22(5):e16540 [FREE Full text] [doi: 10.2196/16540] [Medline: 32452809]
102. Renew JR, Ladlie B, Gorlin A, Long T. The impact of social media on anesthesia resident recruitment. J Educ Perioper Med 2019;21(1):E632 [FREE Full text] [Medline: 31406704]
103. 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]
104. Ruan B, Yilmaz Y, Lu D, Lee M, Chan TM. Defining the digital self: a qualitative study to explore the digital component of professional identity in the health professions. J Med Internet Res 2020 Sep 29;22(9):e21416 [FREE Full text] [doi: 10.2196/21416] [Medline: 32990636]
105. Duymuş TM, Karadeniz H, Şükür E, Atiç R, Zehir S, Azboy İ. Social media and internet usage of orthopaedic surgeons. J Clin Orthop Trauma 2017;8(1):25-30 [FREE Full text] [doi: 10.1016/j.jcot.2016.10.007] [Medline: 28360492]
106. Long X, Qi L, Ou Z, Zu X, Cao Z, Zeng X, et al. Evolving use of social media among Chinese urologists: opportunity or challenge? PLoS One 2017;12(7):e0181895 [FREE Full text] [doi: 10.1371/journal.pone.0181895] [Medline: 28753632]
107. Wang Z, Wang S, Zhang Y, Jiang X. Social media usage and online professionalism among registered nurses: a cross-sectional survey. Int J Nurs Stud 2019 Oct;98:19-26. [doi: 10.1016/j.ijnurstu.2019.06.001] [Medline: 31255853]
108. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 07;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]
109. Henry RK, Pieren JA. The use of social media in dental hygiene programs: a survey of program directors. J Dent Hyg 2014 Aug;88(4):243-249. [Medline: 25134957]
110. Langenfeld SJ, Vargo DJ, Schenarts PJ. Balancing privacy and professionalism: a survey of general surgery program directors on social media and surgical education. J Surg Educ 2016;73(6):e28-e32. [doi: 10.1016/j.jsurg.2016.07.010] [Medline: 27524278]
111. Marnocha S, Marnocha MR, Pilliow T. Unprofessional content posted online among nursing students. Nurse Educ 2015;40(3):119-123. [doi: 10.1097/NNE.0000000000000123] [Medline: 25501656]

112. Al Qarni AM, Al Shehri SZ, Wahab MMA. Perceptions and attitudes of family medicine residency program directors regarding ranking applicants for residency programs using their social media accounts: a national study in Saudi Arabia. J Family Community Med 2019;26(2):133-140 [FREE Full text] [doi: 10.4103/jfcm.JFCM_176_18] [Medline: 31143087]
113. Staud SN, Kearney RC. Social media use behaviors and state dental licensing boards. J Dent Hyg 2019 Jun;93(3):37-43. [Medline: 31182567]
114. Brisson GE, Fisher MJ, LaBelle MW, Kozmic SE. Defining a mismatch: differences in usage of social networking sites between medical students and the faculty who teach them. Teach Learn Med 2015;27(2):208-214. [doi: 10.1080/10401334.2015.1011648] [Medline: 25893945]
115. Duke VJA, Anstey A, Carter S, Gosse N, Hutchens KM, Marsh JA. Social media in nurse education: utilization and e-professionalism. Nurse Educ Today 2017 Oct;57:8-13. [doi: 10.1016/j.nedt.2017.06.009] [Medline: 28683342]
116. Garg M, Pearson DA, Bond MC, Runyon M, Pillow MT, Hopson L, et al. Survey of individual and institutional risk associated with the use of social media. West J Emerg Med 2016 May;17(3):344-349 [FREE Full text] [doi: 10.5811/westjem.2016.2.28451] [Medline: 27330669]
117. Jafarey A, Shekhani S, Azam M, Gill R, Shirazi B, Hassan M, et al. Physicians in cyberspace: finding boundaries. Asian Bioethics Rev 2016;8(4):272-289. [doi: 10.1353/asb.2016.0023]
118. Kitsis EA, Milan FB, Cohen HW, Myers D, Herron P, McEvoy M, et al. Who's misbehaving? Perceptions of unprofessional social media use by medical students and faculty. BMC Med Educ 2016 Feb 18;16:67 [FREE Full text] [doi: 10.1186/s12909-016-0572-x] [Medline: 26887561]
119. Soares W, Shenvi C, Waller N, Johnson R, Hodgson CS. Perceptions of unprofessional social media behavior among emergency medicine physicians. J Grad Med Educ 2017 Feb;9(1):85-89 [FREE Full text] [doi: 10.4300/JGME-D-16-00203.1] [Medline: 28261400]
120. Greysen SR, Johnson D, Kind T, Chretien KC, Gross CP, Young A, et al. Online professionalism investigations by state medical boards: first, do no harm. Ann Intern Med 2013 Jan 15;158(2):124-130. [doi: 10.7326/0003-4819-158-2-201301150-00008] [Medline: 23318312]
121. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
122. Panahi S, Watson J, Partridge H. Social media and physicians: exploring the benefits and challenges. Health Informatics J 2016 Jun;22(2):99-112 [FREE Full text] [doi: 10.1177/1460458214540907] [Medline: 25038200]
123. Markham MJ, Gentile D, Graham DL. Social media for networking, professional development, and patient engagement. Am Soc Clin Oncol Educ Book 2017;37:782-787 [FREE Full text] [doi: 10.1200/EDBK_180077] [Medline: 28561727]
124. Wink DM. Social networking sites. Nurse Educ 2010;35(2):49-51. [doi: 10.1097/NNE.0b013e3181ced776] [Medline: 20173583]
125. Recommendations for: social media usage and maintaining privacy, confidentiality and professionalism. National Council of State Boards of Nursing. URL: https://www.ncsbn.org/NSNA_Social_Media_Recommendations.pdf [accessed 2020-10-22]
126. Mazman SG, Usluel YK. Modeling educational usage of Facebook. Comput Education 2010 Sep;55(2):444-453. [doi: 10.1016/j.compedu.2010.02.008]
127. Bosch TE. Using online social networking for teaching and learning: Facebook use at the University of Cape Town. Communicatio 2009 Nov;35(2):185-200. [doi: 10.1080/02500160903250648]
128. Mason R. Learning technologies for adult continuing education. Stud Continuing Education 2006 Jul;28(2):121-133. [doi: 10.1080/01580370600751039]
129. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res 2016 Jan 04;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
130. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. Acad Med 2015 Jun;90(6):718-725. [doi: 10.1097/ACM.0000000000000700] [Medline: 25785682]
131. Ryan G. International perspectives on social media guidance for nurses: a content analysis. Nurs Manag (Harrow) 2016 Dec 01;23(8):28-35. [doi: 10.7748/nm.2016.e1555] [Medline: 27905234]
132. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. Acad Med 2013 Jun;88(6):893-901. [doi: 10.1097/ACM.0b013e31828ffc23] [Medline: 23619071]
133. Maloney S, Moss A, Ilic D. Social media in health professional education: a student perspective on user levels and prospective applications. Adv Health Sci Educ Theory Pract 2014 Dec;19(5):687-697. [doi: 10.1007/s10459-014-9495-7] [Medline: 24566977]
134. Bahner DP, Adkins E, Patel N, Donley C, Nagel R, Kman NE. How we use social media to supplement a novel curriculum in medical education. Med Teach 2012;34(6):439-444. [doi: 10.3109/0142159X.2012.668245] [Medline: 22449268]
135. Pander T, Pinilla S, Dimitriadis K, Fischer MR. The use of Facebook in medical education--a literature review. GMS Z Med Ausbild 2014;31(3):Doc33. [doi: 10.3205/zma000925] [Medline: 25228935]
136. Prensky M. Digital natives, digital immigrants part 1. On Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]

137. Peck JL. Social media in nursing education: responsible integration for meaningful use. J Nurs Educ 2014 Mar;53(3):164-169. [doi: 10.3928/01484834-20140219-03] [Medline: 24530130]
138. Chauhan B, George R, Coffin J. Social media and you: what every physician needs to know. J Med Pract Manage 2012;28(3):206-209. [Medline: 23373164]
139. Cabrera D. Mayo Clinic includes social media scholarship activities in academic advancement. Medium. 2016. URL: https://tinyurl.com/cm6whmcz [accessed 2021-11-06]
140. Gorrindo T, Groves JE. Web searching for information about physicians. JAMA 2008 Jul 09;300(2):213-215. [doi: 10.1001/jama.2008.44] [Medline: 18612120]
141. Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res 2011 Jan 19;13(1):e8 [FREE Full text] [doi: 10.2196/jmir.1590] [Medline: 21247862]
142. Jent JF, Eaton CK, Merrick MT, Englebert NE, Dandes SK, Chapman AV, et al. The decision to access patient information from a social media site: what would you do? J Adolesc Health 2011 Oct;49(4):414-420 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.004] [Medline: 21939873]
143. Moubarak G, Guiot A, Benhamou Y, Benhamou A, Hariri S. Facebook activity of residents and fellows and its impact on the doctor-patient relationship. J Med Ethics 2011 Feb;37(2):101-104. [doi: 10.1136/jme.2010.036293] [Medline: 21160080]
144. Brown J, Ryan C, Harris A. How doctors view and use social media: a national survey. J Med Internet Res 2014 Dec 02;16(12):e267 [FREE Full text] [doi: 10.2196/jmir.3589] [Medline: 25470407]
145. George DR, Green MJ, Navarro AM, Stazyk KK, Clark MA. Medical student views on the use of Facebook profile screening by residency admissions committees. Postgrad Med J 2014 May;90(1063):251-253. [doi: 10.1136/postgradmedj-2013-132336] [Medline: 24578513]
146. Nagler A, Andolsek K, Rudd M, Sloane R, Musick D, Basnight L. The professionalism disconnect: do entering residents identify yet participate in unprofessional behaviors? BMC Med Educ 2014 Mar 27;14:60 [FREE Full text] [doi: 10.1186/1472-6920-14-60] [Medline: 24674275]
147. Mostaghimi A, Crotty BH. Professionalism in the digital age. Ann Intern Med 2011 Apr 19;154(8):560-562. [doi: 10.7326/0003-4819-154-8-201104190-00008] [Medline: 21502653]
148. Mawdsley A, Schafheutle EI. Using Facebook to support learning and exam preparation in a final-year undergraduate pharmacy clinical therapeutics module. Curr Pharm Teaching Learning 2015 Nov;7(6):869-875. [doi: 10.1016/j.cptl.2015.08.010]
149. Cain J, Scott DR, Tiemeier AM, Akers P, Metzger AH. Social media use by pharmacy faculty: student friending, e-professionalism, and professional use. Curr Pharm Teaching Learning 2013 Jan;5(1):2-8. [doi: 10.1016/j.cptl.2012.09.002]
150. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
151. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Med Educ Online 2010 Sep 15;15:5324 [FREE Full text] [doi: 10.3402/meo.v15i0.5324] [Medline: 20859533]
152. Tran K, Morra D, Lo V, Quan SD, Abrams H, Wu RC. Medical students and personal smartphones in the clinical environment: the impact on confidentiality of personal health information and professionalism. J Med Internet Res 2014 May 22;16(5):e132 [FREE Full text] [doi: 10.2196/jmir.3138] [Medline: 24855046]
153. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 02;127(13):1413-1421. [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]
154. Finn G, Garner J, Sawdon M. 'You're judged all the time!' Students' views on professionalism: a multicentre study. Med Educ 2010 Aug;44(8):814-825. [doi: 10.1111/j.1365-2923.2010.03743.x] [Medline: 20633221]
155. Chretien KC, Goldman EF, Beckman L, Kind T. It's your own risk: medical students' perspectives on online professionalism. Acad Med 2010 Oct;85(10 Suppl):S68-S71. [doi: 10.1097/ACM.0b013e3181ed4778] [Medline: 20881708]
156. Jain A, Petty EM, Jaber RM, Tackett S, Purkiss J, Fitzgerald J, et al. What is appropriate to post on social media? Ratings from students, faculty members and the public. Med Educ 2014 Feb;48(2):157-169. [doi: 10.1111/medu.12282] [Medline: 24528398]
157. White J, Kirwan P, Lai K, Walton J, Ross S. 'Have you seen what is on Facebook?' The use of social networking software by healthcare professions students. BMJ Open 2013;3(7):E003013 [FREE Full text] [doi: 10.1136/bmjopen-2013-003013] [Medline: 23883886]
158. Chretien KC, Greysen SR, Chretien J, Kind T. Online posting of unprofessional content by medical students. JAMA 2009 Sep 23;302(12):1309-1315. [doi: 10.1001/jama.2009.1387] [Medline: 19773566]
159. Hardouin S, Cheng TW, Mitchell EL, Raulli SJ, Jones DW, Siracuse JJ, et al. RETRACTED: Prevalence of unprofessional social media content among young vascular surgeons. J Vasc Surg 2020 Aug;72(2):667-671. [doi: 10.1016/j.jvs.2019.10.069] [Medline: 31882313]
160. Laughney C. What you need to know about the infamous #MedBikini study. Medium. URL: https://medium.com/beingwell/what-you-need-to-know-about-the-infamous-medbikini-study-1d814e1ebe44 [accessed 2021-04-18]

161. Annear S. #MedBikini: here's why health care professionals are posting photos of themselves in bathing suits. The Boston Globe. URL: https://www.bostonglobe.com/2020/07/24/metro/medbikini-heres-why-healthcare-professionals-are-posting-photos-themselves-bathing-suits-social-media/ [accessed 2021-04-18]
162. Hardouin S, Cheng TW, Mitchell EL, Raulli SJ, Jones DW, Siracuse JJ, et al. Retraction notice. J Vasc Surg 2020 Oct;72(4):1514. [doi: 10.1016/j.jvs.2020.08.018] [Medline: 32972597]
163. Pronk SA, Gorter SL, van Luijk SJ, Barnhoorn PC, Binkhorst B, van Mook WNKA. Perception of social media behaviour among medical students, residents and medical specialists. Perspect Med Educ 2021 Aug;10(4):215-221 [FREE Full text] [doi: 10.1007/s40037-021-00660-1] [Medline: 33826108]
164. Meira TM, Prestes J, Gasparello GG, Antelo OM, Pithon MM, Tanaka OM. The effects of images posted to social media by orthodontists on public perception of professional credibility and willingness to become a client. Prog Orthod 2021 Mar 08;22(1):7 [FREE Full text] [doi: 10.1186/s40510-021-00353-9] [Medline: 33682012]
165. Greysen SR, Chretien KC, Kind T, Young A, Gross CP. Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. JAMA 2012 Mar 21;307(11):1141-1142. [doi: 10.1001/jama.2012.330] [Medline: 22436951]
166. Spector N, Kappel DM. Guidelines for using electronic and social media: the regulatory perspective. Online J Issues Nurs 2012 Sep 30;17(3):1 [FREE Full text] [Medline: 23036057]
167. Wissinger CL, Stiegler Z. Using the extended parallel process model to frame e-professionalism instruction in healthcare education. Teach Learn Med 2019;31(3):335-341. [doi: 10.1080/10401334.2018.1528155] [Medline: 31157561]
168. Al-Khalifa KS, Al-Swuailem AS, AlSheikh R, Muazen YY, Al-Khunein YA, Halawany H, et al. The use of social media for professional purposes among dentists in Saudi Arabia. BMC Oral Health 2021 Jan 12;21(1):26 [FREE Full text] [doi: 10.1186/s12903-021-01390-w] [Medline: 33435911]
169. Gay S. School of Medicine: social media guidance for medical students. University of Nottingham. 2019. URL: https://www.nottingham.ac.uk/medicine/documents/medicine-policies/social-media-guidance.pdf [accessed 2020-10-22]
170. Medical student social media and online activity policy. Indiana University School of Medicine. 2021. URL: https://policies.medicine.iu.edu/policies/edu-adm-0007/index.html#history [accessed 2021-11-04]
171. MMS physicians' guide to social media. The Massachusetts Medical Society. 2015. URL: http://www.massmed.org/Governance-and-Leadership/Committees,-Task-Forces-and-Sections/MMS-Physicians--Guide-to-Social-Media-(pdf)/ [accessed 2021-11-04]
172. Social media: practical guidance and best practice. British Medical Association. 2018. URL: https://www.bma.org.uk/media/1852/bma-social-media-practical-guidance-2018.pdf [accessed 2020-10-22]
173. Social media for family physicians: guidelines and resources for success. American Academy of Family Physicians. 2013. URL: https://www.aafp.org/dam/AAFP/documents/about_site/SocialMediaFamPhys.pdf [accessed 2020-10-22]
174. Social media and medical professionalism: a guide to online professionalism for medical practitioners and medical students. Australian Medical Association. 2020 Feb. URL: https://ama.com.au/sites/default/files/Social_Media_and_the_Medical_Profession_FINAL.pdf [accessed 2020-10-22]
175. WMA statement on the professional and ethical use of social media. World Medical Association. 2011 Oct. URL: https://www.wma.net/policies-post/wma-statement-on-the-professional-and-ethical-use-of-social-media/ [accessed 2020-10-22]
176. Kind T. Professional guidelines for social media use: a starting point. AMA J Ethics 2015 May 01;17(5):441-447 [FREE Full text] [doi: 10.1001/journalofethics.2015.17.5.nlit1-1505] [Medline: 25986088]
177. Shore R, Halsey J, Shah K, Crigger B, Douglas SP, AMA Council on Ethical and Judicial Affairs. Report of the AMA Council on Ethical and Judicial Affairs: professionalism in the use of social media. J Clin Ethics 2011;22(2):165-172. [Medline: 21837888]
178. Farnan JM, Snyder Sulmasy L, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627 [FREE Full text] [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
179. Hennessy CM, Smith CF, Greener S, Ferns G. Social media guidelines: a review for health professionals and faculty members. Clin Teach 2019 Oct;16(5):442-447. [doi: 10.1111/tct.13033] [Medline: 31144449]
180. Cain J, Scott DR, Akers P. Pharmacy students' Facebook activity and opinions regarding accountability and e-professionalism. Am J Pharm Educ 2009 Oct 01;73(6):104 [FREE Full text] [doi: 10.5688/aj7306104] [Medline: 19885073]
181. Williams J, Feild C, James K. The effects of a social media policy on pharmacy students' facebook security settings. Am J Pharm Educ 2011 Nov 10;75(9):177 [FREE Full text] [doi: 10.5688/ajpe759177] [Medline: 22171105]
182. Hall M, Hanna L, Huey G. Use and views on social networking sites of pharmacy students in the United kingdom. Am J Pharm Educ 2013 Feb 12;77(1):9 [FREE Full text] [doi: 10.5688/ajpe7719] [Medline: 23459621]
183. Chretien KC, Kind T. Climbing social media in medicine's hierarchy of needs. Acad Med 2014 Oct;89(10):1318-1320 [FREE Full text] [doi: 10.1097/ACM.0000000000000430] [Medline: 25076202]

184. Economides JM, Choi YK, Fan KL, Kanuri AP, Song DH. Are we witnessing a paradigm shift?: A systematic review of social media in residency. Plast Reconstr Surg Glob Open 2019 Aug;7(8):e2288 [FREE Full text] [doi: 10.1097/GOX.0000000000002288] [Medline: 31592016]
185. Parsi K, Elster N. Why can't we be friends? A case-based analysis of ethical issues with social media in health care. AMA J Ethics 2015 Nov 01;17(11):1009-1018 [FREE Full text] [doi: 10.1001/journalofethics.2015.17.11.peer1-1511] [Medline: 26595241]
186. Roy D, Taylor J, Cheston CC, Flickinger TE, Chisolm MS. Social media: portrait of an emerging tool in medical education. Acad Psychiatry 2016 Feb;40(1):136-140. [doi: 10.1007/s40596-014-0278-5] [Medline: 25800704]
187. Alsobayel H. Use of social media for professional development by health care professionals: a cross-sectional web-based survey. JMIR Med Educ 2016 Sep 12;2(2):e15 [FREE Full text] [doi: 10.2196/mededu.6232] [Medline: 27731855]
188. Fenwick T. Social media and medical professionalism: rethinking the debate and the way forward. Acad Med 2014 Oct;89(10):1331-1334 [FREE Full text] [doi: 10.1097/ACM.0000000000000436] [Medline: 25076200]
189. Harrison B, Gill J, Jalali A. Social media etiquette for the modern medical student: a narrative review. Int J Med Students 2014 Dec 24;2(2):64-67. [doi: 10.5195/ijms.2014.86]
190. de Melo Simplício AH. Social media and dentistry: ethical and legal aspects. Dental Press J Orthod 2019;24(6):80-89 [FREE Full text] [doi: 10.1590/2177-6709.24.6.080-089.sar] [Medline: 31994651]

Abbreviations

EM: emergency medicine
FtP: Fitness to Practice
GDC: General Dental Council
HCP: health care professional
HIPAA: Health Insurance Portability and Accountability Act
IRR: interrater reliability
MeSH: Medical Subject Headings
PD: program director
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SM: social media
SNS: social networking site

Edited by R Kukafka; submitted 14.11.20; peer-reviewed by WSY Chan, Y Yilmaz; comments to author 14.01.21; revised version received 19.04.21; accepted 23.09.21; published 17.11.21.

Please cite as:
Vukušić Rukavina T, Viskić J, Machala Poplašen L, Relić D, Marelić M, Jokic D, Sedak K
Dangers and Benefits of Social Media on E-Professionalism of Health Care Professionals: Scoping Review
J Med Internet Res 2021;23(11):e25770
URL: https://www.jmir.org/2021/11/e25770
doi:10.2196/25770
PMID:34662284

Original Paper

Cancer Communication and User Engagement on Chinese Social Media: Content Analysis and Topic Modeling Study

Liang Chen[1], PhD; Pianpian Wang[2], PhD; Xin Ma[3], MA; Xiaohui Wang[4], PhD

[1]School of Journalism and Communication, Tsinghua University, Beijing, China

[2]School of Media and Communication, Shenzhen University, Shenzhen, China

[3]Department of Communication, University of Maryland, Maryland, MD, United States

[4]Department of Media and Communication, City University of Hong Kong, Hong Kong, China

Corresponding Author:
Pianpian Wang, PhD
School of Media and Communication
Shenzhen University
#3883 Baishi Road
Nanshan District
Shenzhen, 518000
China
Phone: 86 15013528703
Email: pian.wang@szu.edu.cn

Abstract

Background: Cancer ranks among the most serious public health challenges worldwide. In China—the world's most populous country—about one-quarter of the population consists of people with cancer. Social media has become an important platform that the Chinese public uses to express opinions.

Objective: We investigated cancer-related discussions on the Chinese social media platform Weibo (Sina Corporation) to identify cancer topics that generate the highest levels of user engagement.

Methods: We conducted topic modeling and regression analyses to analyze and visualize cancer-related messages on Weibo and to examine the relationships between different cancer topics and user engagement (ie, the number of retweets, comments, and likes).

Results: Our results revealed that cancer communication on Weibo has generally focused on the following six topics: social support, cancer treatment, cancer prevention, women's cancers, smoking and skin cancer, and other topics. Discussions about social support and cancer treatment attracted the highest number of users and received the greatest numbers of retweets, comments, and likes.

Conclusions: Our investigation of cancer-related communication on Weibo provides valuable insights into public concerns about cancer and can help guide the development of health campaigns in social media.

(J Med Internet Res 2021;23(11):e26310) doi:10.2196/26310

KEYWORDS

cancer-related information; social media; topic modeling; user engagement; Weibo; cancer

Introduction

Background

Cancer ranks among the most widespread threats to public health worldwide. In China—the world's most populous country—roughly one-quarter of the population consists of people with cancer. According to the most recent statistics, approximately 4,292,000 individuals have been diagnosed with aggressive cancer, and 2,890,000 people in China died from the disease in 2015 [1].

During the past decade, social media has become an important platform that the public uses to express opinions [2]. The number of Chinese people who use social media specifically to exchange cancer-related information has increased dramatically [3,4]. Among the most popular social media sites in China, Weibo (Sina Corporation)—China's equivalent of Twitter—has gradually become the primary channel that scholars use to help

them understand public concerns about health-related issues. However, due to the massive amounts of data on Weibo and other social media platforms, efficiently extracting public opinions from social media data has emerged as a critical challenge for researchers [5].

In response to this challenge, this study was conducted to investigate how China's public has discussed cancer on Weibo by summarizing data on the platform via automatic text analytics. First, we conducted topic modeling to categorize cancer-related messages on Weibo. Second, a regression analysis was performed to examine the relationships between various cancer topics and user engagement. The findings could help health care practitioners with designing cancer prevention programs for responding to public concerns, thereby increasing these programs' effectiveness.

Literature Review

Traditionally, face-to-face communication between physicians and patients is the primary method of obtaining health information from patients with cancer and their families. However, with the advent of information and communication technologies, an increasing number of people have been using social media to obtain, generate, and share various information about cancer [6]. Social media has become the main channel where individuals can discuss potential causes and treatments of the disease as well as express their thoughts and emotions about cancer (ie, those that they cannot otherwise share with relatives and friends due to social stigmatization) [7]. In addition, a large number of health professionals have considered the availability of health information on the internet to be a positive development and have created and disseminated cancer-related information on social media [8]. Weibo, as one of the most popular social media platforms in China, has also provided a space that Chinese people can use to share and discuss various health-related issues. As such, Weibo is not only a major source of web-based health information; it also provides scholars with unprecedented opportunities to gauge the public's perceptions and their understanding of health-related issues in China [9,10].

Scholars have conducted content analyses of social media platforms to examine how the public perceives certain diseases, such as H1N1 and breast cancer [11,12]. In recent years, some studies have examined health-related information on social media by using social computing methods in order to explore the general public's attitudes, thoughts, and feelings toward certain issues in an unobtrusive and comprehensive way [4,13]. For instance, Han et al [14] examined public attitudes and perceptions regarding types of cancer by analyzing cancer-related messages on Twitter via semantic network analysis. They found that breast cancer, lung cancer, and prostate cancer were the three most frequently discussed types of cancer on Twitter. Moreover, discussions about cancer contained positive and negative feelings toward and concerns about cancer.

Although extant research has focused on the number of messages about different types of cancer and the public's sentiment toward cancer on Twitter, the nature of cancer-related messages on Chinese social media is still understudied. To address this gap in the literature, we used text analytics to answer the following research question: what does the Chinese public talk about when discussing cancer on Weibo (research question 1)?

Some scholars have also investigated the influences that a message's content has on user engagement when cancer-related messages are disseminated on social media platforms. Social media engagement is measured in terms of users' interactions with a message. This includes having a conversation with other users through comments and performing actions that demonstrate support, such as liking or retweeting [15,16]. Engagement with a social media message is generally recognized as the most important indicator of popularity and public interest [17]. Some studies have indicated that messages about a specific topic or those that include certain elements generate different levels of user engagement. For example, Wang et al [18] found that messages that describe personal cancer-related experiences resulted in a higher level of engagement (ie, more shares and comments), while cancer prevention–related topics failed to engage social media users. Chen et al [19] also found that social media messages about fear and treatment efficacy attracted more public attention and received a greater number of likes, comments, and retweets. In this study, we investigated the relationships between topics of cancer-related messages and user engagement when such messages are discussed and shared on Weibo. User engagement was measured based on the number of comments, likes, and retweets. Therefore, we proposed the following research question: which cancer topics generate the highest levels of user engagement on Weibo (research question 2)?

Methods

Data Collection

We searched for the keyword *cancer* and terms that represented 25 types of cancer to identify cancer-related messages on Weibo that were posted between June 2015 and June 2016. Instead of mining a full year's worth of data, we randomly selected 7 consecutive weeks and downloaded all tweets that were published during this period by using a Python web crawler. Using consecutive weeks as the sampling frame is one of the most commonly used approaches to obtaining data for content analyses, and this method was confirmed to have relatively good efficiency and accuracy in previous studies [20]. After extracting the data, we formed a data set that contained each message's content, the time of posting, the number of retweets, the number of comments, the number of likes, and the users who posted each message. We excluded advertisements and messages that were not related to cancer from the data set. The final data set consisted of 16,654 cancer-related messages on Weibo.

Prior to the analysis, we cleaned the data by using the standard preprocessing steps developed in previous studies [21,22]. The procedures included converting the words in messages to lowercase words; removing stop words, punctuation, numbers, and nonword characters; and stemming the remaining text. Afterward, we extracted all nouns from the text by using the *JiebaR* package in R (R Foundation for Statistical Computing), since nouns still carry meaning even without context and are useful for identifying topics in the text of messages on social media. Due to the messages' language distribution

characteristics, we expected a vast number of very infrequent words in the vocabulary of a collection [23]. Hence, we also removed all of the words that appeared fewer than 5 times in the data set. This is one step in the standard procedures for processing textual data [24], and this step can enhance an algorithm's performance remarkably and stabilize the stochastic inference of latent Dirichlet allocation (LDA) [25].

Topic Modeling With LDA

Topic modeling algorithms are statistical methods that are used to analyze words in a text to discover the major themes in the text [26]. Topic modeling also enables users to organize and summarize a large number of documents that would be impossible to annotate manually. We used the LDA topic modeling technique, which generated the following two probability distribution outputs: the probability distribution of topics in each document and the probability distribution of terms representing each topic. We determined the number of topics by repeating our analysis with different numbers of topics and comparing the perplexity of each analysis. A lower value of perplexity represents a better model fit [27], and increasing the number of topics generally reduces the perplexity value. When choosing the number of topics, one should consider both the simplicity and interpretability of the textual content [22].

Regression Analysis

A regression analysis was conducted to investigate how the topics discussed in cancer-related Weibo posts influenced user engagement. The unit of analysis was the text of a Weibo post. The dependent variable was engagement, which was operationalized as the number of retweets, number of comments, and number of likes that a message received. The independent variables were the cancer-related topics that were revealed by the topic modeling analysis. Further, this study considered a series of control variables to control the potential influence that message characteristics had on user engagement, including the number of nouns in a message; whether a message contained a URL, a mention of usernames, or a hashtag; and whether a message was news. The independent variables were entered into the regression model according to their assumed causal order. Specifically, control variables were entered in the first block, and the second block included the topics. The probability distributions generated by the LDA analysis revealed the topic data for each message and the probability of a message falling into a particular topic category. The distribution of the number of nouns was left skewed. Thus, a square root transformation was performed prior to the analysis. The number of retweets, comments, and likes followed a power-law distribution. A log transformation was performed prior to the analysis.

Results

Word Frequency Statistics

According to the frequency statistics, the keyword that was mentioned the most often in the data set was *breast cancer* (1465/16,654, 8.80%), followed by *leukemia* (1315/16,654, 7.90%), *esophageal cancer* (1155/16,654, 6.93%), and *lymph cancer* (1121/16,654, 6.73%). Other cancers that were frequently discussed included lung cancer (808/16,654, 4.85%), gastric cancer (761/16,654, 4.57%), cervical cancer (729/16,654, 4.34%), and brain cancer (637/16,654, 3.83%). In summary, discussions about cancer on Weibo focused on the following three cancer categories: digestive cancers (ie, esophageal cancer, gastric cancer, intestinal cancer, gallbladder carcinoma, and liver cancer), urogenital cancers (ie, cervical cancer, metrocarcinoma, endometrial cancer, uterine cancer, bladder cancer, prostate cancer, renal cancer, renal adenocarcinoma, and ovarian cancer), and lymphoid and hematopoietic cancers (ie, leukemia, lymph cancer, and lymphoma). In addition, a large part of the discussions referred to cancer and malignancy in a more general sense.

Topic Modeling via LDA

By conducting topic modeling to analyze the text and comparing the perplexity indices, we found that categorizing topics into 6 topic categories was optimal. Based on our findings, as well as findings from previous cancer-related studies [27], we identified the following six topics.

Topic 1: Social Support

The topic of social support encompassed a wide range of opinions about forms of social support and available help for coping with cancer. The most commonly used keywords that represented social support were *child*, *hope*, *love*, *friend*, *father*, *mother*, *classmate*, *son*, *daughter*, and *family*. In addition, the keyword *leukemia* was also found in discussions about social support, largely because relatives and friends of people with leukemia tend to seek social support on Weibo.

Topic 2: Cancer Treatment

The topic of cancer treatment represented discussions involving the keywords *treatment*, *tumor*, *early stage*, *later stage*, *symptom*, *patient*, *surgery*, *hospital*, *method*, *chemotherapy*, *diagnose*, and *appearance* and other words that immediately suggested the topic of cancer treatment.

Topic 3: Cancer Prevention

The topic of cancer prevention generally concerned issues related to cancer prevention, especially food and eating habits. The most frequently used keywords that represented this topic were *risk*, *food*, *foodstuff*, *research*, *vegetables*, *diet*, *prevention*, *fruit*, *anticancer*, *drinking*, *chance*, *intake*, *soy milk*, *cured food*, *vitamins*, and *fat*. As such, this topic covered not only cancer prevention but also relationships among food, eating habits, and cancer risks.

Topic 4: Women's Cancers

The most frequently used keywords that represented the specific characteristics of women's cancers were *female*, *uterus*, *breast cancer*, *cervical cancer*, *HPV*, *age*, *menstruation*, *fertility*, and *ovarian cancer*. Discussions about this topic included several common types of women's cancers and important related topics, such as the human papilloma virus, menstruation, fertility, and age.

Topic 5: Smoking and Skin Cancer

This topic represented smoking-related cancers and skin cancer in particular. The most frequently used keywords included

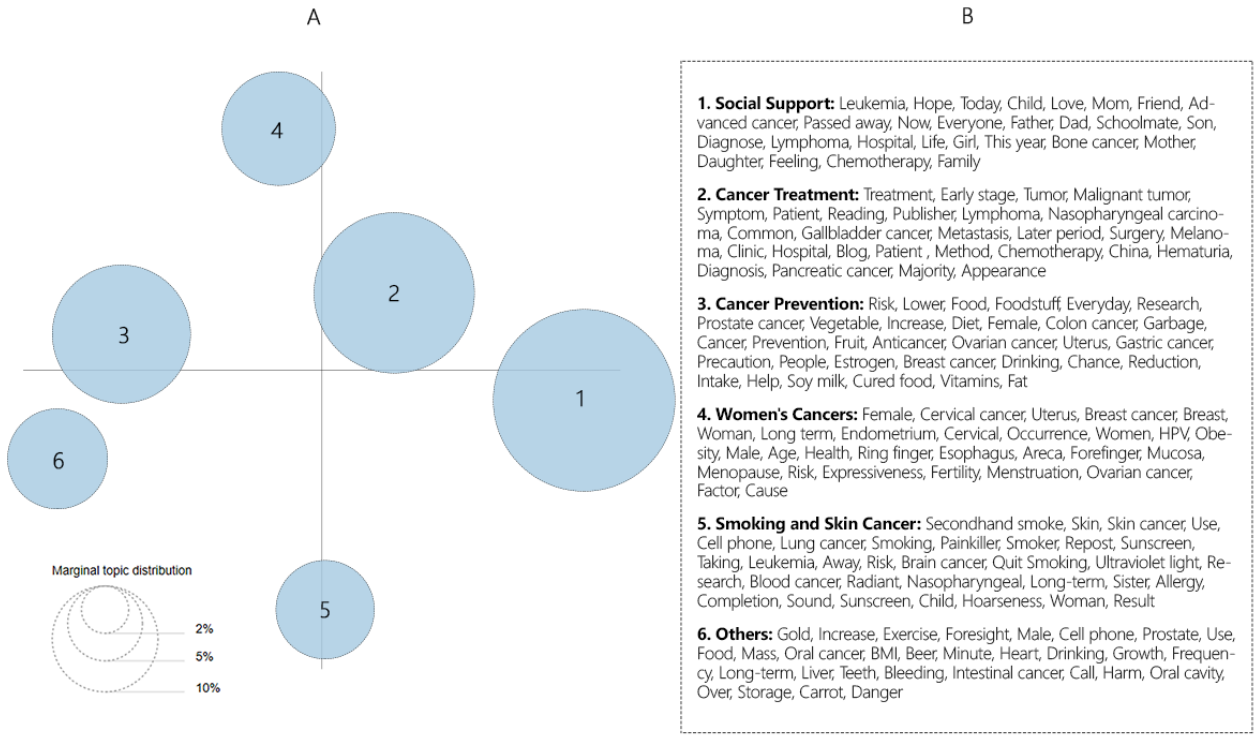
secondhand smoke, *skin cancer*, *lung cancer*, *smoker*, *sunscreen*, *risk*, *quit smoking*, *ultraviolet light*, and *child*.

Topic 6: Others

We classified the sixth topic as "others," since it contained several topics related to cancer prevention, men's cancers, alcohol, and other types of cancer. The most frequently used keywords were *golden rules for cancer prevention*, *exercise*, *cell phone*, *male*, *prostate*, *BMI*, *heart*, *beer*, *drinking*, *oral cancer*, and *intestinal*.

Figure 1 shows the intertopic distance of all topics.

Figure 1. Intertopic distance map. A: LDAvis package and the 30 most common words representing each topic; B: translated from Chinese terms. The size of the circles indicates the topic proportions for all text. HPV: Human Papilloma Virus.



Descriptive Analysis of Engagement

Table 1 presents the descriptive statistics and intercorrelations of the three measures of engagement. The number of retweets for all of the messages varied from 0 to 8033 (mean 10.73, SD 169.54). The number of comments ranged from 0 to 10,897 (mean 6.38, SD 142.08). The number of likes ranged from 0 to 48,963 (mean 20.21, SD 516.83). The distribution of all of the engagement measures followed a long-tail distribution; a few messages generated a high level of engagement, while most of the others languished in obscurity. Moreover, the correlation tests showed a strong correlation between the number of retweets and comments, while the binary relationships between these two popularity indices and the number of likes were moderate. This indicated that the three measures were highly related and presented unique aspects of user engagement.

Table 1. Descriptive analysis and correlation matrix of user engagement measures.

Popularity measures	Number of retweets[a]	Number of comments[b]	Number of likes[c]
Number of retweets			
r	1	0.732[d]	0.414[d]
P value	—[e]	<.001	<.001
Number of comments			
r	0.732[d]	1	0.484[d]
P value	<.001	—	<.001
Number of likes			
r	0.414[d]	0.484[d]	1
P value	<.001	<.001	—

[a]The mean number of retweets was 10.73 (SD 169.54).

[b]The mean number of comments was 6.38 (SD 142.08).

[c]The mean number of likes was 20.21 (SD 516.83).

[d]Significant at the *P*<.001 level.

[e]Not applicable.

Regression Analysis Results

Table 2 presents the results from the ordinary least squares regression analysis for predicting engagement. All factors were significantly associated with the number of retweets, and social support was the primary factor related to the number of retweets. In addition, social support, cancer treatment, women's cancers, and smoking and skin cancer were positively related to the number of comments. The topic of cancer prevention was not significantly associated with the number of comments on messages (*P*=.09). It means that the coefficient of it (.016) is not significant. Besides, all factors were positively associated with the number of likes, except for the inclusion of URLs (*P*=.07) and the cancer prevention topic (*P*=.06).

The results of the regression analysis indicated that among all five cancer topics, discussions about social support had the greatest effect on user engagement indices, that is, the number of retweets (β=.26; *P*<.001), comments (β=.36; *P*<.001), and likes (β=.34; *P*<.001). In other words, social support was the most popular cancer topic on Weibo. Messages that contained discussions about social support were more likely to attract users' attention and stimulate engagement on social media. Further, the cancer treatment and smoking and skin cancer topics were also strong predictors of user engagement, which suggests that these topics attract a lot of public attention.

Table 2. Regression results based on user engagement measures as the criteria (N=16,651).

Variables	Number of retweets[a], β[b]	Number of comments[c], β[b]	Number of likes[d], β[b]
Intercept	.001	.001	0
Control variables			
Noun count	.118[e]	–.005[e]	.056[e]
Inclusion of a URL	.066[e]	–.050[e]	.014[e]
Inclusion of a username mention	.044[e]	.028[e]	.023[f]
Inclusion of a hashtag	.114[e]	.057[e]	.126[e]
The message was news	.144[e]	.042[e]	.108[e]
Topics			
Social support	.262[e]	.361[e]	.337[e]
Cancer treatment	.125[e]	.097[e]	.116[e]
Cancer prevention	.033[e]	.016[e]	.023[e]
Women's cancers	.054[e]	.038[e]	.059[e]
Smoking and skin cancer	.110[e]	.084[e]	.123[e]

[a]The R^2 value for the number of retweets is 0.066 (P<.001).

[b]Indicates the β weights or standardized regression weights.

[c]The R^2 value for the number of comments is 0.089 (P<.001).

[d]The R^2 value for the number of likes is 0.075 (P<.001).

[e]Significant at the P<.001 level.

[f]P=.002

Discussion

Principal Findings

Our study made an initial attempt at conducting a systematic textual analysis of cancer-related messages on Weibo. By using word frequency statistics and conducting topic modeling and regression analyses, we identified the major topics of public discussions about cancer on Chinese social media and examined the relationship between different cancer topics and user engagement.

The word frequency statistics indicated that the cancer that is discussed the most often on Weibo is breast cancer, followed by leukemia, esophageal cancer, and lymphoma. This contradicts the incidence and mortality rates of cancer in China [1], where lung cancer, stomach cancer, and liver cancer are the most prevalent cancers, followed by breast cancer. A plausible explanation for this is that the survival rate of patients with breast cancer is greater than the survival rates of most patients with other types of cancer. Consequently, many patients with breast cancer and their families might spend a longer time expressing their opinions and seeking relevant information on Weibo [28,29]. In addition, according to Weibo's annual report, 68.8% of Weibo users are aged 18-35 years, 77.8% of users have a high level of education, and users in high-income areas account for a relatively large proportion of users [30]. In other words, a majority of Weibo users are young people with high socioeconomic statuses. However, those with a higher risk of cancer and mortality are typically older adults and people with low socioeconomic statuses. As such, we found a difference between the cancers discussed on Weibo and the cancers with the highest mortality rates in China.

We used automatic text analytics to examine cancer-related discussions on Weibo. The discussions were categorized into the following six topics: social support, cancer treatment, cancer prevention, women's cancers, smoking and skin cancer, and other topics. Our topic categories summarize the main topics that are discussed by the public when they talk about cancer, and they are statistically supported by a large amount of data. In particular, the results revealed that people mainly seek and provide social support messages about cancer on Weibo. Moreover, our regression results confirm that social support is the most popular cancer topic on Weibo. Messages about social support received a larger number of retweets, comments, and likes than those about other topics. These results align with findings from previous studies [29,31] and can possibly be explained by the Chinese collectivist culture, in which people's decision-making is usually interdependent. When developing medical responses to health problems, Chinese people tend to rely on advice and support from health care professionals and people with similar health conditions [32].

Our results also suggest that many public discussions about cancer on Chinese social media focus on cancer treatment. In addition, messages about cancer treatment received a large number of retweets, comments, and likes. This suggests that most Weibo users pay more attention to (ie, retweet, comment on, and like) messages that discuss cancer symptoms and

treatments and pay less attention to other topics, such as cancer prevention. However, a closer look at the discussions indicates that most messages on Weibo exaggerate the effects that diet has on cancer treatment. For instance, the following message captures this tendency quite well:

> *[Onion wine is effective in treating cancer] Tu Ge is a 40-year-old man. After diagnosed with cancer, he drank onion wine every day. Then, the cancer cells are getting smaller.*

Although such messages on Weibo attract a lot of attention and generate a lot of discussion, such inaccurate statements can misguide individuals' responses to cancer [33]. Further, while health-related misinformation is a long-standing area of research [34], studies on the effects of nonmalicious misinformation remain scarce. Future studies could investigate specific categories of misinformation about cancer, especially the difference between malicious and nonmalicious cancer-related misinformation.

Interestingly, our results show that while people do not discuss smoking and skin cancer very often on Weibo, messages about smoking-related topics are more likely to generate a high level of user engagement. One possible explanation for this could be that unlike other cancer topics, such as cancer treatment and prevention, that attract people living with cancer, smoking is a common issue in China. As a result, smokers and their relatives and friends may be interested in sharing and engaging in smoking-related discussions. Thus, smoking-related messages generate a higher level of user engagement on Weibo.

Our findings provide several insights. First, the thematic framework that we discovered via topic modeling could complement and verify the traditional categorization of cancer-related topics. The categorization of public discussions also sheds light on the public's knowledge and awareness of cancer-related issues [35]. Second, by combining automatic text analytics and regression analyses, we provided an approach to summarizing and gaining insights from a large amount of user-generated content. In terms of practicality, this means that the topic modeling of cancer-related content on Weibo can be conducted to monitor the public's attention to cancer in real time. Third, the regression analysis showed that the social support and cancer treatment topics attracted the most public attention. These findings could help relevant agencies and health care professionals to understand social media users' needs and suggest strategies for designing effective health campaigns and promotions.

Limitations and Suggestions for Future Research

Our study has several limitations. First, because data on social media change greatly over time, the messages posted at the time of this study may not fully reflect the current situation. Additionally, since our analysis did not account for the time of posting, the dynamic nature of the public's attention to cancer-related topics remains unknown. Especially since the State Council of China issued the guideline on Internet Plus Healthcare in 2018, the health information industry in China has witnessed rapid growth and structural changes. Therefore, on the basis of this study, we call for further studies on this topic to compare discussions conducted at different time points to gain insights into how health-related discussions on social media will evolve over time.

Second, it is possible that the cancer-related posts of users who are more popular on social media platforms attracted more attention. Thus, the number of followers that an individual Weibo user has could be a strong confounder of the number of retweets, comments, and likes. However, this factor was not taken into consideration in this study due to the lack of techniques and resources. Future studies can conduct hierarchical regression analyses to investigate this issue.

Finally, we analyzed data from Weibo only. Therefore, any generalization of our results should take the characteristics of Weibo users into consideration [36]. Future research should be conducted to analyze data on other social media platforms and cross-verify our results.

Conflicts of Interest

None declared.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
2. Chen L, Yang X. Nature and effectiveness of online social support for intercultural adaptation of mainland Chinese international students. International Journal of Communication 2015;9:2161-2181.
3. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, 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]
4. Song H, Omori K, Kim J, Tenzek KE, Hawkins JM, Lin WY, et al. Trusting social media as a source of health information: Online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]
5. Luo Y. The internet and agenda setting in China: The influence of online public opinion on media coverage and government policy. International Journal of Communication 2014;8:24.
6. Hong Y, Peña-Purcell NC, Ory MG. Outcomes of online support and resources for cancer survivors: a systematic literature review. Patient Educ Couns 2012 Mar;86(3):288-296. [doi: 10.1016/j.pec.2011.06.014] [Medline: 21798685]

7. Sharf BF. Communicating breast cancer on-line: support and empowerment on the internet. Women Health 1997;26(1):65-84. [doi: 10.1300/J013v26n01_05] [Medline: 9311100]
8. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
9. Rauchfleisch A, Schäfer MS. Multiple public spheres of Weibo: a typology of forms and potentials of online public spheres in China. Inf Commun Soc 2014 Jul 24;18(2):139-155. [doi: 10.1080/1369118x.2014.940364]
10. Stockmann D, Luo T. Which social media facilitate online public opinion in China? Problems of Post-Communism 2017 Mar 30;64(3-4):189-202 [FREE Full text] [doi: 10.1080/10758216.2017.1289818]
11. 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]
12. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb 04;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
13. Zhou Y, Moy P. Parsing framing processes: The interplay between online public opinion and media coverage. Journal of Communication 2007 Mar;57(1):79-98. [doi: 10.1111/j.0021-9916.2007.00330.x]
14. Han G, Zhu D. Talking about cancer on Twitter: Health semantics and social media. Chinese Journal of Journalism and Communication 2017;39(4):44-62.
15. Drula G. Social and online media research – data, metrics and methods. Review of Applied Socio-Economic Research 2012;3(1):77-86 [FREE Full text]
16. Leander M. What is a good engagement rate on a Facebook page? Here is a benchmark for you. Michael Leander. URL: https://www.cyberclick.net/numericalblogen/what-is-a-good-engagement-rate-for-the-different-social-networks [accessed 2019-07-03]
17. Niciporuc T. Comparative analysis of the engagement rate on Facebook and Google Plus social networks. In: Proceedings of International Academic Conferences. 2014 Presented at: 13th International Academic Conferences; September 15-18, 2014; Antibes, France.
18. Wang X, Chen L, Shi J, Peng TQ. What makes cancer information viral on social media? Comput Human Behav 2019 Apr;93:149-156. [doi: 10.1016/j.chb.2018.12.024]
19. Chen L, Yang X, Fu L, Liu X, Yuan C. Using the extended parallel process model to examine the nature and impact of breast cancer prevention information on mobile-based social media: Content analysis. JMIR Mhealth Uhealth 2019 Jun 24;7(6):e13987 [FREE Full text] [doi: 10.2196/13987] [Medline: 31237239]
20. Riffe D, Aust CF, Lacy SR. The effectiveness of random, consecutive day and constructed week sampling in newspaper content analysis. Journal Mass Commun Q 1993 Mar 01;70(1):133-139. [doi: 10.1177/107769909307000115]
21. Lee TY, Bradlow ET. Automated marketing research using online customer reviews. J Mark Res 2011 Oct;48(5):881-894. [doi: 10.1509/jmkr.48.5.881]
22. Mankad S, Han HS, Goh J, Gavirneni S. Understanding online hotel reviews through automated text analysis. Serv Sci 2016 May 05;8(2):124-138. [doi: 10.1287/serv.2016.0126]
23. Manning C, Schütze H. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT; 2003.
24. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, et al. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Commun Methods Meas 2018 Feb 16;12(2-3):93-118. [doi: 10.1080/19312458.2018.1430754]
25. Denny MJ, Spirling A. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. Polit Anal 2018 Mar 19;26(2):168-189. [doi: 10.1017/pan.2017.44]
26. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
27. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
28. Hu W, Adey P. A scientific creativity test for secondary school students. Int J Sci Educ 2002;24(4):389-403. [doi: 10.1080/09500690110098912]
29. Shi J, Chen L. Social support on Weibo for people living with HIV/AIDS in China: a quantitative content analysis. Chinese Journal of Communication 2014 Jul 24;7(3):285-298. [doi: 10.1080/17544750.2014.926954]
30. Weibo user development report of 2016. Weibo. URL: https://data.weibo.com/report/reportDetail?id=346 [accessed 2021-05-03]
31. Han X, Li B, Qu J, Zhu Q. Weibo friends with benefits for people live with HIV/AIDS? The implications of Weibo use for enacted social support, perceived social support and health outcomes. Soc Sci Med 2018 Aug;211:157-163. [doi: 10.1016/j.socscimed.2018.06.016] [Medline: 29940485]
32. Kim MS, Smith DH, Yueguo G. Medical decision making and Chinese patients' self-construals. Health Commun 2009 Dec 10;11(3):249-260. [doi: 10.1207/s15327027hc110307]
33. Chen L, Wang X, Peng TQ. Nature and diffusion of gynecologic cancer-related misinformation on social media: Analysis of tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]
34. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]

35. Shi J, Poorisat T, Salmon CT. The use of social Networking Sites (SNSs) in Health Communication Campaigns: Review and Recommendations. Health Commun 2018 Jan;33(1):49-56. [doi: 10.1080/10410236.2016.1242035] [Medline: 27858464]
36. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? 2010 Apr Presented at: WWW '10: The 19th International World Wide Web Conference; April 26-30, 2010; Raleigh, North Carolina, USA p. 591-600. [doi: 10.1145/1772690.1772751]

Abbreviations

LDA: latent Dirichlet allocation

Edited by R Kukafka; submitted 06.12.20; peer-reviewed by W Pian, S Zhang, L Liu, DY Kim; comments to author 15.03.21; revised version received 10.05.21; accepted 08.10.21; published 10.11.21.

Please cite as:
Chen L, Wang P, Ma X, Wang X
Cancer Communication and User Engagement on Chinese Social Media: Content Analysis and Topic Modeling Study
J Med Internet Res 2021;23(11):e26310
URL: https://www.jmir.org/2021/11/e26310
doi:10.2196/26310
PMID:34757320

Original Paper

Causal Effects of Alcohol-Related Facebook Posts on Drinking Behavior: Longitudinal Experimental Study

Hanneke Hendriks[1,2], PhD; Wouter de Nooy[2], PhD; Winifred A Gebhardt[3], PhD; Bas van den Putte[2], Prof Dr

[1]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

[2]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[3]Department of Health, Medical and Neuropsychology, Leiden University, Leiden, Netherlands

Corresponding Author:
Hanneke Hendriks, PhD
Behavioural Science Institute
Radboud University
Thomas van Aquinostraat 4
Nijmegen, 6525 GD
Netherlands
Phone: 31 24 3612372
Email: hanneke.hendriks@ru.nl

Abstract

Background: Adolescents and young adults frequently post alcohol-related content (ie, alcoholposts) on social media. This is problematic because both social norms theory and social learning theory suggest that viewing alcoholposts of peers could increase drinking behavior. It is therefore paramount to understand the effects of exposure to alcoholposts on viewers.

Objective: This study aimed to investigate the causal effects of exposure to alcoholposts on alcohol consumption by using a rigorous design.

Methods: We conducted a 6-week longitudinal study during which alcoholposts were measured by a newly developed app that copied Facebook posts shared by participants (n=281) to a new social media environment. In addition, daily questionnaires assessed alcohol use. Effects of natural alcoholposts (ie, posted by the participants) were assessed in phase 1, and effects of experimental posts (ie, posted by fake participants) were explored in phase 2.

Results: Results showed that natural alcoholposts increased the occurrence and quantity of drinking the following day. That is, exposure to a single additional alcoholpost increased the log odds of drinking the next day by 0.27 (b=.27, credible interval [CI] .18 to .35). Furthermore, the number of natural alcoholposts had a positive (predictive) effect on the number of glasses drunk the next day (b=.21, CI .14 to .29). In phase 2 when experimental posts were also present, these effects decreased. Experimental posts themselves had hardly any effects.

Conclusions: This study illustrates clear and direct effects of exposure to alcoholposts on next-day alcohol consumption and suggests that alcoholposts represent an important societal problem that interventions need to address.

(J Med Internet Res 2021;23(11):e28237) doi:10.2196/28237

KEYWORDS

social media; social networking site (SNS); alcohol-related posts; alcoholposts; alcohol consumption

Introduction

Interpersonal Communication About Alcohol Use

Adolescent alcohol abuse is related to severe accidents, brain damage, and future alcohol addiction [1,2]. Studies in offline, face-to-face contexts have shown that interpersonal communication about alcohol influences alcohol consumption [3,4]. However, as a result of significant changes in the interpersonal communication and media landscape, communication today often takes place online [5]. In particular, adolescents and young adults frequently visit social networking sites (SNSs) such as Facebook and Instagram and often post alcohol-related content on these sites [6,7]. Examples of alcohol-related SNS content (henceforth: alcoholposts) are party pictures posted on Facebook in which groups of people are holding alcoholic drinks or Instagram photos in which a close-up of a cocktail is shown [8]. Given the prevalence of alcoholposts,

it is paramount to understand the effects of exposure to alcoholposts on their viewers.

Some recent studies suggest that alcoholposts are related to increased alcohol consumption. Unfortunately, however, most of these studies do not allow for strong conclusions about causality because they use cross-sectional designs [6,9-11], making it impossible to conclude whether posting or seeing more alcoholposts leads people to consume more alcohol or vice versa. Furthermore, previous studies frequently rely on retrospective self-reported social media use [8,12-14], which can be problematic because people may have difficulty remembering what they have encountered on social media and at what exact time, especially if the inquired recollection overarches a longer period (eg, in the past year). Last, not many studies have experimentally investigated effects of alcoholposts in a realistic social media environment. Thus, to ascertain the effects of alcoholposts on alcohol use, there is an urgent need for longitudinal studies that combine daily measurements of alcohol consumption with daily objective measurements of alcohol-related social media content and that experimentally study the effects of alcoholposts on drinking behavior.

Alcohol-Related Content on Social Media

The rise of social media introduced new ways for young people to communicate with each other. A multitude of studies examining platforms such as Facebook, Instagram, and YouTube show that young people post many alcoholposts. Although estimated percentages of young people posting alcoholposts vary between 36% to 96%, studies agree that alcohol is frequently visible on SNS [6,7,15]. Most adolescents and young adults indicate that they post alcoholposts because they want to entertain others or celebrate and share nice moments with friends [16]. Furthermore, it has been suggested that posting alcoholposts is something that people are generally not consciously aware of, as it often happens without deliberate intent [17].

Content analyses have drawn several important conclusions about the nature of alcoholposts: (1) they mostly portray alcohol use as a normal part of life (eg, pictures of dinners or parties), (2) they almost always involve a positive context (eg, laughing faces), and (3) they often contain a social component (eg, showing groups of people) [8,18,19]. All 3 aspects are problematic. The first 2 because they imply that alcoholposts express to viewers that alcohol use is positive and normal, while at the same time the negative consequences of alcohol use (eg, drunk people embarrassing themselves) are ignored. The third aspect is also alarming, because a vast body of research consistently confirms that social norms are very powerful in influencing behavior [20-23]. Thus, viewing alcoholposts in which many people are drinking alcohol is likely to have a strong influence on behavior. Given these characteristics and the presence of alcoholposts on social media, it is vital to better understand how viewing such posts influences drinking behavior.

Why Viewing Alcoholposts May Make People Drink

The tenets of both social norms theory [24] and social learning theory [25] would predict that seeing alcoholposts strongly increases drinking behavior among viewers. Social learning theory and later social cognitive theory [26] posit that behaviors can be learned from observing and imitating others. Moreover, it is suggested that especially seeing behavior rewarded or punished can lead to learning effects and behavior changes. As argued, alcoholposts are often positive and social (eg, showing laughing people being present at fun events). These aspects of alcoholposts may make alcohol use seem rewarding, leading viewers to increase their drinking behaviors.

Social norms theory similarly suggests that behavior is influenced by people's perceptions of how others think and act. These perceived norms are more important for behavior than objective norms (ie, what others actually do), while they are likely to be based on misperceptions of how others behave [27]. In the context of alcohol use, it has been shown that people often mistakenly think that others drink more alcohol than they do themselves (ie, pluralistic ignorance [28]). This false belief may be reinforced by exposure to alcoholposts in which seemingly everyone is engaging in drinking behaviors. Such a belief may particularly increase drinking behaviors in viewers who want to fit in and behave in line with existing norms (see also Beullens and Vandenbosch [29]).

Studies on the Relationship Between Alcoholposts and Alcohol Consumption

Thus, in line with both social norms theory and social learning theory, viewing alcoholposts of peers on social media is expected to increase drinking behavior. Several recent studies have addressed this relationship; however, some important aspects of these studies limit the implications of this previous research.

Cross-sectional Studies

Several cross-sectional studies suggest that exposure to alcoholposts is related to drinking behaviors [9,10]. For example, Ranney et al [30] linked state-wide alcohol tweets with emergency care visits. Geusens and Beullens [6] observed that self-reported alcoholposts on social media were related to increased self-reports of alcohol abuse, and Thompson and Romo [11] found that self-reported alcoholposts were associated with alcohol-induced problems. However, in these studies questionnaires measured both seeing alcoholposts in a previous period and alcohol use in the previous period. Therefore, it is not clear whether alcoholposts predict drinking or are a consequence of (ie, simply reflect) it.

Longitudinal Studies

A handful of recent studies have used a longitudinal design to address the relationship between exposure to alcoholposts and alcohol use. For example, Tucker and colleagues [15] showed that greater self-reported exposure to alcoholposts in 7th grade predicted drinking in 8th grade, controlled for drinking at 7th grade. Erevik and colleagues [31] showed that especially self-reported disclosure (instead of exposure) of alcoholposts predicted alcohol use 1 year later. Furthermore, in a study by Boyle et al [12], self-reported exposure to alcoholposts during the first 6 weeks of college predicted alcohol use 6 months later, controlled for previous alcohol use. However, these longitudinal studies all measured exposure to alcoholposts by using

self-report (eg, questions such as "How often during the past 3 months did you see pictures on an SNS showing or talking about someone who is drunk?" [9]). This is problematic, since evidence suggests that it is very hard to correctly recall media exposure (ie, recall bias), especially if it covers a longer time period in the past [13,14]. The problem of self-reported media exposure is not unique to this context and poses a challenge for researchers in many other related fields [32].

Furthermore, these previous studies used a long time span between exposure to alcoholposts and alcohol use (eg, 6 or 12 months), whereas a short time span (eg, 1 day) would provide useful insights into the direct effects of viewing an alcoholpost on drinking shortly afterward. That is, although longitudinal designs with longer time spans may provide information on long-term effects of alcoholposts, it is unclear what occurs exactly in between waves, making the direct and short-term effects of alcoholposts unknown. Thus, although a few valuable longitudinal studies exist, limitations in their designs curtail our understanding of the direct causal effects of alcoholposts on alcohol use. In this study, we applied a longitudinal design in which we measured exposure to alcoholposts objectively and included daily measurements of exposure to alcoholposts as well as alcohol use.

Experimental Studies

Not many experimental studies exist that address the effect of exposure to alcoholposts on alcohol use. One experimental study by Alhabash et al [33] manipulated various aspects of alcohol marketing posts and studied subsequent effects (see also Alhabash et al [34]). In line with this, another experimental study by Noel [35] on alcohol marketing manipulated types of ads as well as comments and studied effects on purchase intentions. Although very valuable, these studies focused on alcohol marketing and not on user-generated content. Furthermore, these studies did not use a realistic immersive social media environment reflecting real-life conditions. We argue it is essential to use an experimental design to study the effects of alcoholposts in order to be able to make inferences about causality. That is, if participants are randomly assigned to different alcoholpost conditions, then any differences between groups with regard to drinking behavior can be attributed to the exposure to the specific posts within the experimental conditions. Therefore, to solidify our claims about the causal effects of alcoholposts, we also used an experimental design in which the effects of experimentally manipulated exposure to alcoholposts were assessed. We did this in a realistic social media environment closely resembling real life.

Our Study

In sum, although previous studies have provided valuable insights into alcohol-related social media use, to provide a clear answer about the causal effects of exposure to alcohol content on social media, we used a longitudinal study that combined objectively measured daily exposure to alcoholposts with daily measurements of alcohol consumption (ie, whether [occurrence] and how much [quantity] people drink). Because previous research has shown that existing alcoholposts are mainly positive and social and social norms theory, social learning theory, and empirical evidence suggest that seeing alcoholposts leads to increased drinking behavior, we expected the following:

- H1: Exposure to natural alcoholposts increases the occurrence and quantity of alcohol consumption.

Furthermore, in a second phase of our study we experimentally investigated the effects of alcoholposts. More specifically, existing alcoholposts were observed in the first phase (ie, natural posts), and additional alcoholposts were experimentally manipulated during the second phase (ie, experimental posts). The experimental posts differed in terms of valence (negative versus positive about alcohol) and the degree to whether they were social (ie, showed people), and we investigated how these experimental posts influenced daily measures of alcohol consumption. Because previous research [21,36] has shown that the activation of positive alcohol associations (eg, as a consequence of positive alcoholposts) and the observation of other people drinking alcohol (eg, as a consequence of social alcoholposts) can increase alcohol consumption, we expected the following:

- H2a: Exposure to positive experimental alcoholposts increases the occurrence and quantity of alcohol consumption, whereas negative experimental alcoholposts decrease the occurrence and quantity of alcohol consumption.
- H2b: Exposure to social experimental alcoholposts has a stronger influence on the occurrence and quantity of alcohol consumption than exposure to nonsocial alcoholposts.

In this study, we focused on college students, based on a previous study highlighting that college students (aged 18 to 30 years) posted far more alcoholposts than high school students (aged 12 to 18 years) [17]. Furthermore, we chose to focus on Facebook because this was the most popular social media channel among our target group (college students in the Netherlands) at the time the study was conducted. That is, 89% of Dutch people aged 20 to 39 years used Facebook, in comparison with 46% who used Instagram and 32% who used Snapchat [37]. Among those aged 15 to 19 years, Facebook use was also very high (72%), although Instagram (73%) and Snapchat (72%) were starting to get more users. Furthermore, several studies conducted in this context showed that alcoholposts were common on Facebook. For example, Van Hoof et al [38] showed that 99% of college student Facebook profiles contained alcohol references. Therefore, we used Facebook to study effects of alcohol content.

Methods

To obtain daily exposure measurements of alcoholposts, we developed a social media app that copied participants' Facebook posts to a new and realistic social media environment (more information below). Effects were assessed on daily measurements of alcohol consumption.

Participants

Participants were all students and participated in groups (ie, they were asked to sign up as a group: friends, classmates, colleagues). Most groups were friends or classmates. The reason

for recruiting in groups is that we wanted participants to know some but not all of the other participants. That is, we wanted them to see posts of people who were familiar to them (as is normally the case on social media), but we also wanted them to see posts of people unfamiliar to them. That way, we could add experimental posts of fake unfamiliar participants to the study without participants' awareness. In all analyses we controlled for group (ie, we did multilevel analyses).

The baseline survey was answered by 306 participants. During the course of 6 weeks, 25 participants never answered the daily survey and were therefore omitted from analyses. Therefore, 281 participants, who were part of 49 groups, were included in the analyses (208 women, 73 men, mean age [SD 1.90] 20.53 years, range 17 to 30 years). All participants were Dutch. In the Netherlands, the minimum legal age to purchase alcoholic beverages is 18 years.

There were 49 groups in total. Group sizes ranged from 2 to 18 participants. Most groups consisted of 4 to 5 people. That is, 7 (14%) groups had 2 to 3 participants, 25 groups (51%) had 4 to 5 participants, 8 groups (16%) had 6 to 7 participants, 4 groups (8%) had 8 to 9 participants, and 5 groups (10%) had 10 to 18 participants. At the beginning of the study, there were 722 existing Facebook friendships within these groups (out of 922 possible friendships within groups). Across groups, there were 653 friendships (out of 38,418 possible friendships across groups). On average, a participant had a Facebook friendship with 4.6 participants from another group.

Design

The study used a longitudinal design with 43 measurements (1 baseline survey and 42 daily measurements). Participants first completed a baseline survey and starting 1 week later, they were followed for 6 weeks (ie, 42 days in total) during which alcohol-related social media use and alcohol consumption were measured on a daily basis.

The study had 2 phases: during the first 3 weeks, posts were merely observed (ie, natural posts), and during the next 3 weeks, fake participants (ie, profiles made by the experimenter) posted additional alcoholposts (ie, experimental posts). In order to test the influence of different types of experimental posts, we manipulated 2 aspects of the experimental posts. That is, based on studies that show that alcoholposts are usually positive and social [8,19], we manipulated whether the post had a negative or positive context and also whether the post was social or not social (ie, no people visible). Groups were randomly assigned to 1 of the 4 conditions in this 2 (negative versus positive) × 2 (not social versus social) between-subjects design. During phase 2, participants only saw experimental posts within the same condition (eg, only positive social experimental posts).

Procedure

Participants were recruited at university campuses and student buildings and were asked to participate in groups (minimum size of 4 people). At the beginning of the baseline survey, participants were informed about what their participation in the study entailed (ie, completing daily questionnaires, engaging with the SNS app, and giving access to their Facebook posts) after which they provided their informed consent.

To prevent participants from correctly guessing the purpose of the study, they were told that they would be involved in 2 separate study parts (ie, 1 focusing on the questionnaire, and 1 focusing on Facebook using the SNS app). Next, they answered questions concerning demographics and expected covariates (ie, gender, study year, habitual frequency of alcohol use, and habitual quantity of alcohol use). At the end of the baseline survey, participants received instructions on how to download, install, and use the SNS app.

A week thereafter the study started. Participants took part in the study for 6 weeks. Every day, they had 2 tasks: answer a short questionnaire (measuring alcohol use on the previous day) and visit and engage with the SNS app. Filler questions (about exercising and snacking behaviors) were added to the questionnaire to obscure the real purpose of the study. To help remind participants, we sent push messages with a link to the questionnaire via the SNS app every day at 9 AM. Furthermore, participants could click on the SNS app, view posts, and engage (ie, like and/or comment) with posts (more information on the SNS app is described later). Participants were told that it was very important to answer the questionnaire and visit the app every day. We monitored whether participants opened and engaged with the SNS tool (time was not monitored). When participants did not log in, they would receive a reminder to engage with the tool. In the beginning, this was done for everyone on a daily basis. After a few weeks, ensuring that participants were engaging actively, this was done by randomly checking activity levels of participants.

As stated, the study had 2 phases: during the first 3 weeks (days 1 to 21), posts were merely observed. That is, only real posts (ie, natural posts) of the participants were visible in the app. During the next 3 weeks (days 22 to 43), fake participants (ie, profiles made by the experimenter) posted additional alcoholposts (ie, experimental posts; based on the 4 between-subject conditions).

After 43 days, participants answered the final survey after which they were debriefed and rewarded (€30 [US $35] per participant). All participants were extensively debriefed, especially on the fact that there were (positive) fake experimental posts in the second phase and that they should be aware of the negative consequences of alcohol use. At any time, participants were allowed to withdraw their participation, also after reading the debrief (and what the study was really about). The study underwent extensive ethical screening and received ethical approval by the University of Amsterdam's ethical board (2018-PC-8731).

Materials

SNS App

To measure social media use, an app was developed for this study by the software company Akyla. After participants downloaded and accepted the terms of the app, it was able to copy all posts that participants posted on Facebook from that moment on. These posts were subsequently posted in the app, which strongly resembled a Facebook environment. See Figure 1 for a visual representation of the timeline. In this app, participants were able to see their and other participants' posts

and engage (ie, like and/or comment) with them in a similar way as can be done on Facebook. Using this app had several advantages over using Facebook directly: (1) we were absolutely sure what posts and in what order participants were exposed to in the app (with current Facebook algorithms, this is not transparent and individual-specific), (2) we only showed participants posts of other participants (and not of friends of friends) and we only collected participants' posts and not posts of their friends who did not participate in this study, thereby decreasing privacy concerns, and (3) we were able to add experimental posts to this SNS context in a realistic way.

Figure 1. Visual representation of the timeline of the social networking site tool.

Natural Posts

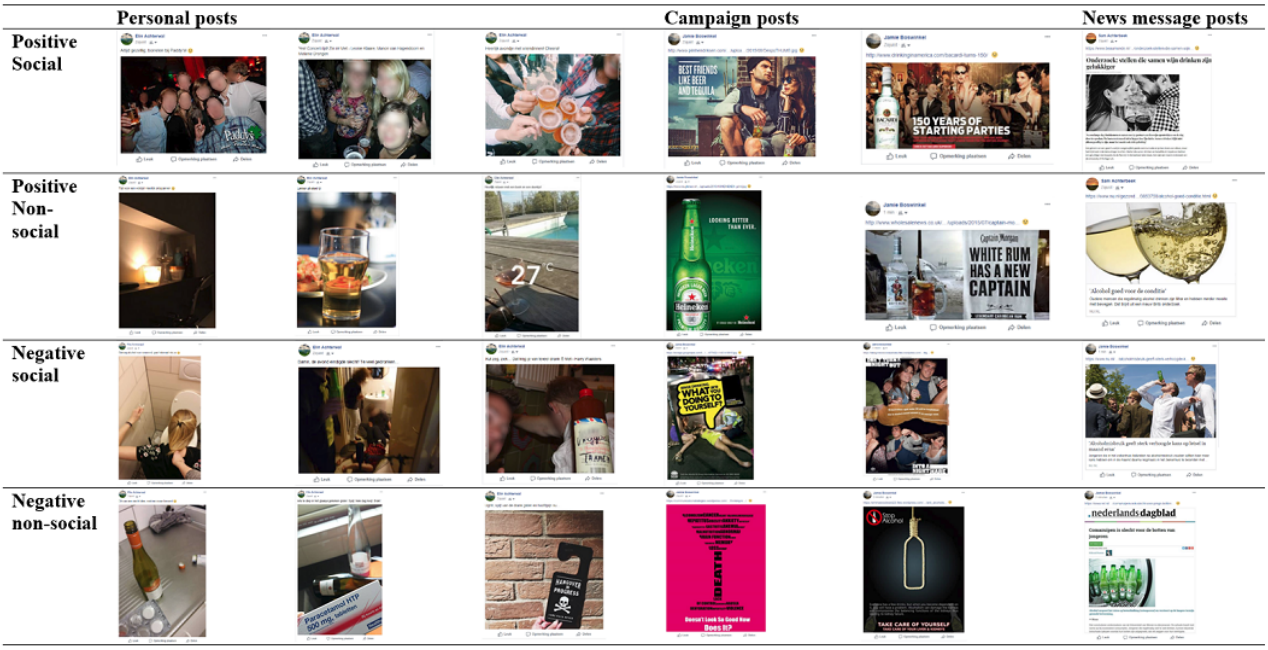
All posts posted in the SNS app were automatically stored for the experimenter in an Excel (Microsoft Corp) file. This Excel file was coded in order to determine whether posts were alcoholposts based on a coding book by Hendriks et al [8]. If a post clearly showed alcohol in the picture or clearly referred to alcohol in the header, it was coded as an alcoholpost. If this was the case, it was also coded whether the context of alcohol was positive (eg, showing laughing people or positive consequences of alcohol use [having fun]), negative (eg, showing frowning people or negative consequences of alcohol use [a hangover]), or neutral (ie, when it was not clearly positive or negative), and whether the alcoholpost was social or not (showing people versus no people visible).

Experimental Posts

There were 24 experimental posts in total. Participants saw 6 experimental alcoholposts within their allocated condition distributed over a period of 3 weeks. All posts fit the condition; however, we used different types of posts to make the manipulation less obvious and provide more variation. That is,

we used 3 personal posts, 2 campaign posts, and 1 news post. The personal posts were similar to the majority of alcoholposts previously reported in literature [8] and showed personal photos of experiences (eg, a night out). The campaign posts reflected posts by professional organizations, either being existing alcohol commercials (in case of the positive conditions) or existing antialcohol campaigns (in case of the negative conditions). The news posts were existing news messages about, depending on the condition, the negative or positive effects of alcohol use. The 24 experimental posts were based on an extensive pilot study (n=41, 29 women, 12 men, mean age [SD 3.23] 22.90 years, range 18 to 36 years), in which 54 posts were evaluated. Pilot participants evaluated the posts on a Likert scale from 1=very negative to 7=very positive by answering questions related to valence ("I think this post is very negative/very positive about alcohol" and "I find it likely that the person posting this is very negative/very positive about alcohol") and social aspects ("How many people did you see/tagged in the post?" and "Do you consider this to be a 'social' post?" and "Did the post include an individual or social activity?"). Of these posts, the most clearly negative/positive and social/nonsocial posts were chosen. See Figure 2 for all experimental posts used and Multimedia Appendix 1 for translations.

Figure 2. All experimental posts used during phase 2.



Measures

Daily Alcohol Consumption

Alcohol consumption was measured in the short daily questionnaire by addressing occurrence (whether people drink) and quantity (how much they drink). Occurrence was measured by the question "Did you drink alcohol yesterday?" (no/yes), and quantity was measured by "How many alcoholic drinks did you consume?" (mean 1.57 [SD 3.27] glasses, range 0 to 50 glasses).

Filler Questions

In the daily questionnaire, filler questions were asked about exercising and snacking behavior to make the focus on alcohol less obvious.

Covariates

The covariates habitual frequency of alcohol use (ie, "How often do you normally consume alcohol?" 8-point scale ranging from "I never drink" to "once a day or more often"), habitual quantity of alcohol use (ie, "On a drinking day, how many alcoholic drinks do you usually have?" 10-point scale ranging from "1" to "10 or more"), gender, and study year were measured in the baseline survey.

Analyses

Multilevel Modeling

We tested multilevel models with 3 levels (ie, daily alcohol use reports [level 1] nested within participants [level 2] nested within groups [level 3]). To examine the effects of alcoholposts on alcohol occurrence, a logistic regression was conducted, and to investigate effects on alcohol quantity, a negative binomial model was tested. We used the Hamiltonian Monte Carlo estimation as implemented in the R package rstanarm [39]. We used a Bayesian approach (Multimedia Appendix 2) to data analysis [40], enabling us to draw conclusions about the probability that a parameter is in a particular range (ie, credible interval). The reader may use this interval for testing null hypothesis significance by checking if the interval contains zero, but we prefer to interpret the intervals because null hypothesis significance testing is highly contested [41,42].

Random Effects

Alcohol drinking is usually a habit linked to particular days in the week [43]. In our sample, we observed clear differences between the weekdays in terms of the number of participants who drank alcohol. In addition, we may expect that students have their own individual habitual drinking days. A separately

estimated multilevel model with days nested within participants suggested that participants varied in terms of the weekdays on which they drink alcohol. The fact that participants had different preferred drinking days may compromise the causal interpretation of effects of alcoholposts on drinking probability and drinking quantity. That is, alcoholposts are likely to show alcohol use that occurred on the posting day or the preceding day. Imagine that participants have different typical drinking days: for example, participant 1 tends to drink on Thursdays and participant 2 drinks on Fridays. Participant 1's Thursday drinking may result in an alcoholpost that precedes the Friday drinking of participant 2; however, the alcoholpost does not cause a higher probability or quantity of drinking for this habitual Friday drinker. For this reason, it is important to control for each participant's individual drinking probability and quantity on each day of the week. In multilevel modeling terms, we included varying (ie, random) effects of the day of the week at the participant level. Effects of exposure to alcoholposts represent changes in the occurrence or quantity of alcohol drinking in comparison to what we normally expect for a participant on this day of the week. In other words, exposure effects show if participants drink more or more often than they normally do on this day of the week.

Random Intercepts

Furthermore, we found that there was variation in average alcohol use and quantity across groups and participants. Some groups drink more often or more glasses of alcohol than other groups. The same applies to participants, where individual drinking differences were visible. Therefore, for both the participant and group, we used random intercepts.

Covariates

The models controlled for the day of the week and for 4 characteristics of the participants: gender (female: no=0), study year, habitual frequency of alcohol use, and habitual quantity of alcohol use (reported during the baseline survey).

Results

Descriptives

Across the 6 weeks of the study, the response rate to complete the daily questionnaire among the remaining 281 participants was 80.8%. Although participants were stimulated to use the SNS tool daily, some days they did not engage with the app. On average, participants used the app on 24 out of 43 days.

There were 547 posts in total. In phase 1, 271 participants shared a post (approximately 1 post per participant), and in phase 2, 194 participants posted something (0.7 post per participant). A total of 39 posts (15 posts in phase 1 and 24 posts in phase 2) were natural alcoholposts, which were posted on 27 separate days. These alcoholposts were posted by 14 participants in phase 1 and 22 participants in phase 2. Focusing on exposure to alcoholposts, the mean number of alcoholposts seen on the previous day was 8.5 (phase 1) and 13.8 (phase 2). There were 7 participants in phase 1 (2.4%) and 15 participants (5.3%) in phase 2 who never viewed an alcoholpost on the previous day. Based on these numbers, we can conclude that almost all participants were exposed to alcoholposts and that this happened frequently.

Coding of these posts revealed that 26 showed a positive context, 11 were neutral, and 2 showed a negative context. Furthermore, 34 of these posts were social (ie, showing people) and 5 were nonsocial (no people visible).

We had exposure measurements for all days except the first day of the observation period. This left us with 8794 alcohol reports (level 1) submitted on 41 days by 281 participants (level 2) who were assigned to 49 groups (level 3). All cases had a valid score on whether the participant drank alcohol; 5 cases had a missing or invalid score for the number of glasses of alcohol consumed and were dropped from the analysis of alcohol quantity consumption.

Hypotheses Testing

Effects of Natural Alcoholposts

We assumed that natural alcoholposts may affect the occurrence of drinking alcohol on the next day and the quantity of alcohol drunk on the next day (H1). The number of alcoholposts created on the day preceding the day with reported alcohol use is our indicator of exposure to alcoholposts. In our model, we tested the effects of natural alcoholposts for the entire period (phase 1 and phase 2); however, we added phase as a moderator so that we could determine the effects of natural posts for each phase separately and also formally test the interaction effect.

Phase 1

We started our analyses by looking at phase 1 (ie, the phase without experimental posts). Results showed that in phase 1, the number of natural alcoholposts had a positive (predictive) effect on the probability of drinking the following day. Exposure to a single additional alcoholpost increased the log odds of drinking the next day by 0.27 (b=.27, credible interval [CI] .18 to .35). This means that, for example, seeing 1 instead of no alcoholpost the day before increased the chance of drinking alcohol from 40% to 47%. Seeing 4 instead of no alcoholposts on the preceding day increased the chance of a drinking day to 66%. (This example is based on drinking on a Saturday for a female, first-year student with average scores on other predictors).

Furthermore, not only occurrence but also quantity of alcohol use was predicted by alcoholpost exposure. That is, the number of natural posts also had a positive (predictive) effect on the number of glasses drunk (b=.21, CI .14 to .29). Using the same example as before, this means that seeing 1 instead of no alcoholpost the day before increased the number of alcohol beverages consumed from 1.64 glasses to 2.03 glasses. Seeing 4 instead of no alcoholposts on the preceding day further increased the number of alcohol beverages consumed to 3.86 glasses. Therefore, H1 is supported for phase 1.

Phase 2

In phase 2, experimental posts were added to the app. As these additional posts may have affected the influence of natural posts, we estimated whether the influence of natural posts depended on the phase (ie, an interaction effect). Indeed, the analyses

showed that there is an interaction effect when predicting the probability of drinking (b=–.35, CI –.48 to –.22) and when predicting the number of glasses drunk (b=–.27, CI –.38 to –.15), suggesting that the effects of natural posts depended on the phase. Indeed, although we found a positive effect of natural alcoholposts on alcohol use in phase 1, we saw no convincing positive effects of natural alcoholposts in phase 2 on occurrence (b=–.08, CI –.17 to 0.1) or quantity (b=–.05, CI –.13 to .03). See Figure 3 for all credible intervals and Figure 4 for an illustration of this interaction effect.

Figure 3. Predicted effects of natural alcoholposts per experimental phase and their 95% and 50% credible intervals for the model predicting occurrence and the quantity of alcohol consumed.

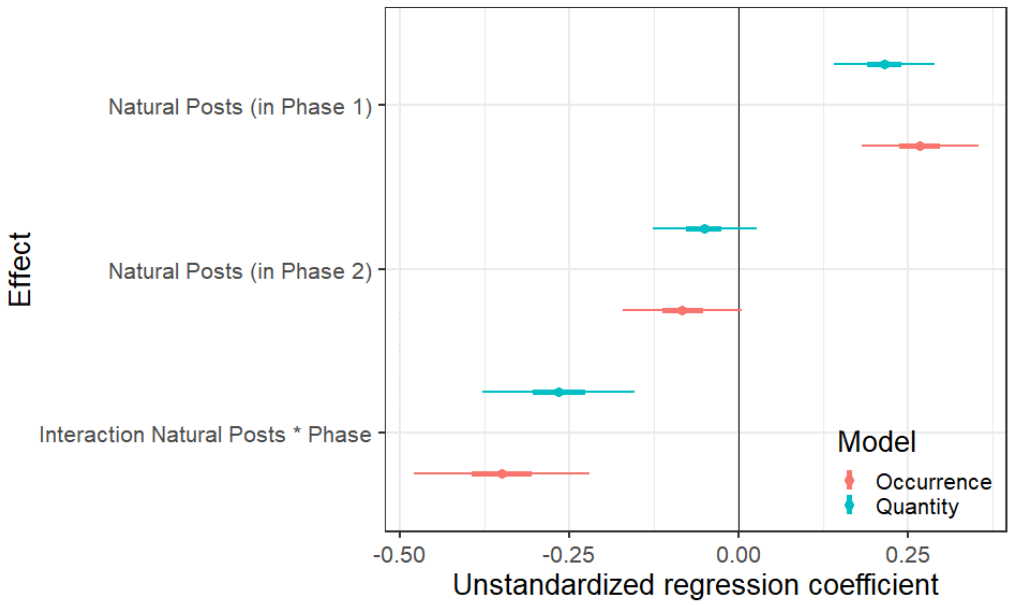
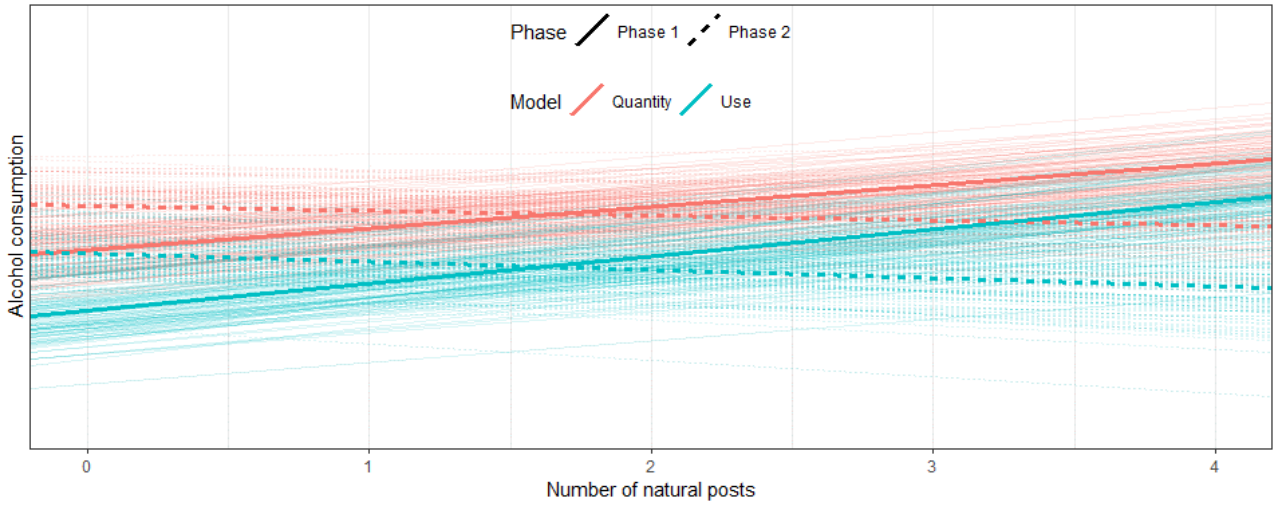


Figure 4. Regression lines for the effects of natural posts on alcohol consumption (occurrence and quantity) in phase 1 and phase 2. The different slopes of the bold and dashed lines illustrate the interaction effect.



Effects of Experimental Posts

H2a predicted that positive experimental alcoholposts lead to increased alcohol occurrence and quantity, whereas negative experimental alcoholposts lead to lower alcohol occurrence and quantity. Results strongly suggest that positive experimental posts have small positive unstandardized regression weights when predicting occurrence (b=.06, CI –.02 to .14, posterior probability of a positive effect is .933), but we are less sure of a positive effect on quantity (b=.02, CI –.05 to .09, posterior probability of a positive effect is .710). Although this suggests that positive experimental posts are more likely to have a positive than negative effect on the occurrence of alcohol use, we are not sufficiently sure about an effect because these credible intervals contain zero.

Looking at negative posts, the effects become even more uncertain and their directions are contradictory. That is, negative experimental posts have small positive unstandardized regression weights when predicting occurrence (b=.02, CI –.06 to .10) and small negative rather than positive unstandardized regression weights when predicting quantity (b=–.02, CI –.10 to .05). These results do not allow us to draw a conclusion about the effects of negative experimental posts. We therefore find little support for H2a.

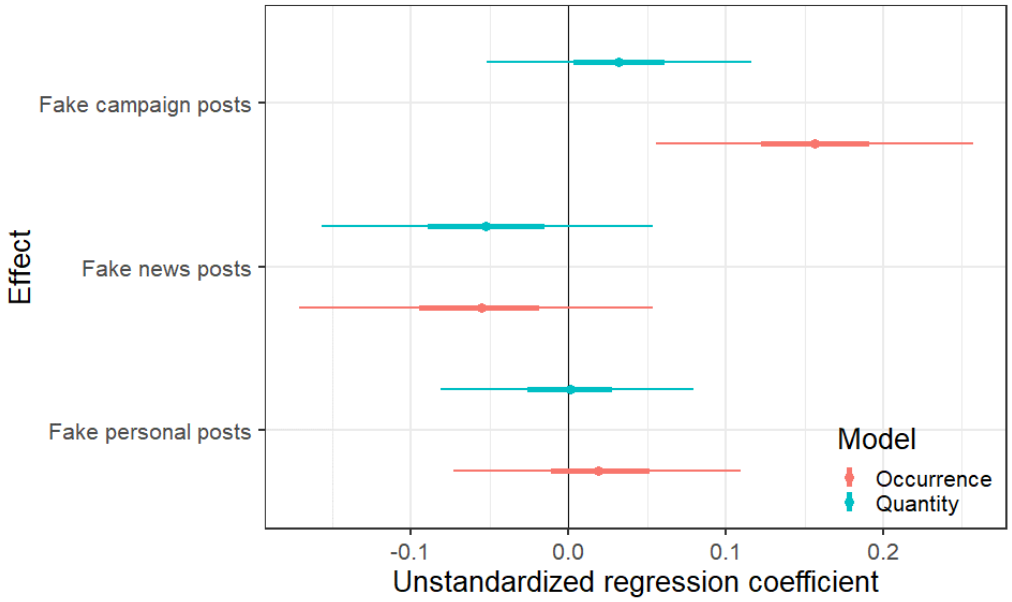
H2b predicted that social experimental alcoholposts had stronger effects on occurrence and quantity than nonsocial posts. Results showed that social experimental posts have small positive unstandardized regression weights when predicting occurrence (b=.05, CI –.03 to .14); however, this was not the case when predicting quantity (b=.00, CI –.08 to .08). Furthermore, nonsocial alcoholposts have small regression weights when predicting alcohol occurrence (b=.02, CI –.06 to .11) and quantity (b=.00, CI –.07 to .07). Although this suggests that social experimental posts are more likely to have a positive than negative effect on alcohol use and these effects seem larger than those from nonsocial posts, we are not sufficiently sure about the effects of social posts because these credible intervals contain zero. Thus, we do not see strong support for H2b.

Type of Posts

As shown in Methods, we used 3 types of experimental alcoholposts (ie, personal posts, campaign messages, and news messages). It is possible that some of these posts are more influential than others, potentially explaining the varying effects of experimental posts described above. Therefore, we exploratively tested whether these types of experimental posts have different effects.

As can be seen in Figure 5, personal or news posts did not hold clear negative or positive effects on alcohol occurrence (b personal=.02 , CI personal –.07 to .11; b news=–.06, CI news –.17 to .05) or quantity (b personal=.00, CI personal –.08 to .08; b news=–.05, CI news –.16 to .05). However, experimental campaign posts (ie, both proalcohol commercials and antialcohol campaigns) had a positive effect on alcohol occurrence (b=.16, CI .06 to .26), although not on quantity (b=.03, CI –.05 to .12). An additional campaign post increases the odds of a drinking day on the next day by 16%. Using the same example as before but now focusing on the second phase, this means that seeing 1 experimental campaign post instead of no experimental alcoholposts on the preceding day increased the chance of drinking alcohol from 53.9% to 57.8%, and seeing 3 campaign posts increased the chance of a drinking day to 65.2%. See Figure 5 for the credible intervals. Please see Multimedia Appendix 3 for additional tables outlining our results, including the influence of control variables.

Figure 5. Predicted effects of experimental campaign, news, and personal alcoholposts and their credible intervals for both models predicting occurrence and quantity of alcohol consumed.



Discussion

Principal Findings

Given that alcoholposts are often present on social media and have potentially undesirable effects on alcohol use, the purpose of this study was to investigate the causal effects of exposure to alcoholposts on alcohol consumption. We conducted a longitudinal study that combined daily measurements of alcohol consumption with objectively measured daily measurements of alcohol-related social media content and that in a second phase also experimentally studied the effects of alcoholposts on alcohol use. Our analyses provide 3 main findings: (1) alcoholposts increase the occurrence and quantity of alcohol consumption on the next day, (2) these effects of alcoholposts disappeared in the second phase of the study when the experimental alcoholposts were posted, and (3) the experimental alcoholposts had hardly any effect on drinking behavior.

The first finding was that exposure to an alcoholpost increased the chance of drinking alcohol as well as the number of alcoholic beverages consumed that following day. Thereby, this study suggests a direct causal effect of exposure to alcoholposts on proximal (next day) alcohol consumption. Although several studies have explored this relationship in a cross-sectional [6,9-11] or longitudinal way [eg, 12], limitations in these designs have restricted the conclusions that could be to drawn on the causal direct effects of alcoholposts. By using daily alcohol consumption measures and daily objective measures of alcoholposts, this study shows that if young people encounter alcoholposts in their social media environment, this increases the chance that and how much they drink the next day.

The second finding was that the effect of exposure to alcoholposts depends on the phase of the study. Whereas in phase 1 the effects of natural alcoholposts on alcohol use were very clear, there were hardly any effects of these alcoholposts visible in phase 2. A potential explanation for this may be that the addition of the experimental posts (6 for each condition) influenced the impact of the natural alcoholposts. Previous research confirms that alcoholposts are positive and social [18,19]. By adding negative and nonsocial posts, we have provided a more diverse alcohol-related social media environment that is not necessarily all proalcohol, which might have dampened the undesirable effects of natural alcoholposts. If this is indeed the case, this provides important ideas for interventions, as this suggests that adding antialcohol content to a social media environment decreases the undesirable impact of alcoholposts. Future research is necessary to investigate whether this indeed is the case.

The third main finding was that exposure to the experimental alcoholposts had almost no effect on drinking behavior. That is, negative experimental posts did not decrease alcohol use, positive alcoholposts did not increase alcohol use, and neither did social posts differ in effects from nonsocial posts. This was quite surprising, as the experimental posts were based on existing alcoholposts and subjected to an extensive pilot study. At first glance, a possible explanation why natural posts had more effects than experimental posts is related to familiarity. That is, the natural posts in our study were posted by real individuals (and experimental posts were posted by fake individuals), making it more likely that the posters were known by other participants. However, we would like to highlight that participants were part of 49 groups, and they were familiar with on average only 4.6 participants from other groups (see Methods). Thus, natural alcoholposts very often were placed by strangers as well. Therefore, we do not think that familiarity can fully explain the differences between experimental and natural posts.

An alternative explanation for why we found almost no effects of experimental posts may be because some of the alcoholposts we used (eg, news posts about alcohol) were not very common in a real social media environment. However, we did find an effect of 1 type of experimental alcoholpost: campaigns. Experimental campaign posts increased whether people drink the next day (but not how much they drink). Interestingly, however, campaigns increased the chance of a drinking day regardless of whether this campaign post was positive about alcohol (ie, an alcohol commercial) or negative about alcohol (ie, an antialcohol campaign). A possible explanation may be that seeing alcohol on social media can serve as a prime and may trigger existing alcohol-related associations that are positive (eg, alcohol is fun [36]), even if alcohol is portrayed negatively. Another explanation may be that antialcohol campaigns increased drinking because of psychological reactance [44]. That is, when individuals feel threatened in their freedom (eg, when a campaign suggests that they should drink less), this may cause reactance against the message. These explanations for this undesirable effect need to be tested in future studies because this may suggest that using antialcohol campaigns on social media may not be a wise strategy.

The purpose of this study was to provide insight into the causal effects of alcoholposts on alcohol use. Because our findings confirm effects of natural alcoholposts but show hardly any effects of experimental posts, we can be relatively sure about order effects but still need more research to be fully certain of the causality of effects. That is, by showing that alcoholposts predict next day drinking and by controlling for personal drinking rates, we show that there are direct acute relations between seeing an alcoholpost and drinking (instead of the other way around). However, we cannot exclude the possibility that third variables exist that relate to both alcoholposts and drinking (eg, alcohol-related events). We therefore recommend additional experimental research manipulating alcoholposts to provide even more clarity on the causality of effects.

Practical Implications

The most important finding of this study is that exposure to alcoholposts increases whether and how much college students drink. Given the abundance of alcohol-related content on social media [6,9], this is a worrisome conclusion. Previous studies have shown high percentages (ie, between 36% and 96%) of young people having alcoholposts on their profile. Our study shows a different perspective on this percentage: that is, we counted 36 respondents (13%) who posted at least 1 alcoholpost, a percentage that is a much lower than those mentioned in previous literature. A potential explanation is that our study is unique in using a short timespan to study the direct effects of alcoholposts. Previous studies have often coded posts in a period of a year or by even coding entire profiles existing of many years. It is therefore not surprising that the latter strategy would yield more alcoholposts than the former. If our study would have focused on a period of 1 year instead of 6 weeks, the percentages found would probably be more in line with previous research.

Although the number of alcoholposts was relatively low in our study, the 39 alcoholposts reflect a large number of people who are exposed to the alcoholposts. On the 43 days of this study, there were 39 instances in which participants saw an alcoholpost. As stated, 1 alcoholpost can already increase the occurrence and quantity of drinking, meaning that exposure to 39 alcoholposts can have a big impact on drinking behavior. We believe this is the crux of the problem of alcoholposts: a single alcoholpost on social media may have enormous reach and simultaneously affect hundreds of people. This problematic impact of alcoholposts becomes even worse if the person posting the alcoholpost is popular and has a large number of friends or followers (ie, is a social influencer [45]), thereby highlighting the need to address this urgent societal issue.

The question then is how to tackle this problem? Although this was not the purpose of the study, we propose 2 potential ways to approach this issue: decrease the posting of alcoholposts and decrease the unhealthy effects of alcoholposts. Many strategies might or might not work in this regard, and future studies need to explore these intervention ideas. To address the first issue, one might need to make young people aware that alcoholposts can pose a real problem and have a negative impact. Also, one could highlight that other people (eg, future employers or parents) might negatively evaluate posters of alcoholposts,

or—as people are often not consciously aware that alcohol is visible in their posts—one could implement automatic warnings on social media when people are about to post an alcoholpost that state "You are about to share a post in which alcohol is visible. Are you sure you want to do that?"

To address the second issue, one could make the social media environment more heterogeneous by adding negative alcohol content. However, based on our finding regarding experimental campaign posts, one should be mindful of the type of negative alcohol content chosen for this purpose. Stimulating peers to also post negative alcohol experiences (eg, hangover posts) might be a possibility. It could also be an idea to stimulate negative comments to alcoholposts or have people withdraw their tags from alcoholposts. Doing so might decrease the normative beliefs in young viewers that alcohol is normal and positive [27-29].

An alternative approach would be to illustrate the "fake" nature of alcoholposts (eg, by showing an alcoholpost with the caption "What you think happened?" next to which an alcohol photo is shown in which a person is lying drunk on the ground with the caption "What happened after"). Future research is needed to test whether these ideas have desirable public health outcomes.

Limitations

Although our study design had several strengths, some limitations should also be noted. First, even though our study used objective social media measurements, alcohol use was measured through self-report. The reason this was done was that measuring alcohol consumption in objective ways (eg, through breathalyzers or observations [46]) for 42 consecutive days would be very difficult to implement in practice. However, to increase the reliability of the alcohol reports, a push message was sent each morning asking about alcohol consumption on the previous day, thereby keeping the length of time between the actual behavior and recollection to a minimum. Although it has been argued that self-reported alcohol consumption measures can be reliable and valid [47], especially if the recall covers a short period in time, future research should try to replicate our findings using more objective measures of alcohol use.

A second limitation might be that our study was relatively intensive by asking participants for daily participation in the app and questionnaire over a period of 6 weeks. This may have potentially led to a decrease in engagement at the end of this period. This could also be a potential explanation why there were hardly any effects in phase 2. Although we had no visible dropout at this time and people still logged in daily to the app during the last 3 weeks, we cannot be sure they were as engaged with the app as they were during the first 3 weeks. Potentially, participants paid less attention to the posts in the app (including the alcoholposts), thereby decreasing their potential impact. This might also explain the limited effects of the experimental posts because these only occurred during the last 3 weeks. Future studies could take into account measures of engagement to address such explanations.

Third, our study focused on the influence of alcohol posts on Facebook because this was the most popular platform among our target group and alcohol posts were common on Facebook. However, in recent years, other social media platforms (eg, Snapchat and Instagram) have gained popularity, especially among adolescent users [48]. Although we expect the effects of a single alcoholpost described in this paper to be visible in other social media contexts as well, it might be the case that the effects are even more pronounced on Instagram or Snapchat. On Instagram, pictures tend to have higher quality and are made more attractive by adding filters [16], potentially leading to more positive and appealing alcohol pictures and possible stronger effects. On Snapchat, posts can be shared privately or only appear for a short amount of time. This could potentially lead to more extreme posts being shared (eg, drunken pictures), with possibly stronger effects on drinking behavior. Future research is therefore needed to investigate how our findings on the effects of alcoholposts translate to other platforms.

Fourth, another limitation is that the SNS app only included posts from college students. Therefore, it is possible that the posts were more homogenous than they would be on a real Facebook timeline, which may also include posts from, for example, family members. However, we did see that the posts in the tool covered many different topics aside from alcohol. Furthermore, in real life, young people mostly interact with their peers on social media (and generally tend to avoid their parents), and Facebook's algorithm ensures that they will especially see posts by like-minded individuals [49]. Nevertheless, although we think that the SNS app resembled real life to a sufficient degree, it might be improved by also including posts from diverse individuals leading to a more heterogenous social media environment.

Conclusion

This study shows a clear and direct effect of exposure to alcoholposts on next-day alcohol consumption. Seeing a natural alcoholpost increases whether young people drink the following day as well as the number of drinks they consume. Although these effects were less visible in the second phase of the study when experimental alcoholposts were also present, these findings suggest that alcoholposts represent an important societal problem that future interventions need to address. Furthermore, the finding that campaigns increased the chance of a drinking day regardless of whether the campaign post was positive (ie, an alcohol commercial) or negative (ie, an antialcohol campaign) about alcohol is relevant for campaign planners and should be further explored in future research.

Acknowledgments

This work was funded by Veni grant 451-15-020 from the Netherlands Organization for Scientific Research awarded to the first author.

Authors' Contributions

HH was responsible for conceptualization, study design, data collection, data analysis, and writing the manuscript. WdN was responsible for data analysis and writing the manuscript. BvdP and WAG were responsible for conceptualization, study design, and writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Translation of posts in Figure 2.
[DOCX File , 16 KB - jmir_v23i11e28237_app1.docx]

Multimedia Appendix 2
Additional information on Bayesian approach.
[DOCX File , 12 KB - jmir_v23i11e28237_app2.docx]

Multimedia Appendix 3
Additional tables.
[DOCX File , 41 KB - jmir_v23i11e28237_app3.docx]

References

1. Bräker AB, Soellner R. Alcohol drinking cultures of European adolescents. Eur J Public Health 2016 Aug;26(4):581-586. [doi: 10.1093/eurpub/ckw033] [Medline: 27037330]
2. Perkins HW. Surveying the damage: a review of research on consequences of alcohol misuse in college populations. J Stud Alcohol Suppl 2002 Mar(14):91-100. [doi: 10.15288/jsas.2002.s14.91] [Medline: 12022733]
3. Hendriks H, van den Putte B, de Bruijn GJ. Changing the conversation: the influence of emotions on conversational valence and alcohol consumption. Prev Sci 2014 Oct;15(5):684-693. [doi: 10.1007/s11121-013-0418-2] [Medline: 23812888]
4. Scholz C, Doré BP, Cooper N, Falk EB. Neural valuation of antidrinking campaigns and risky peer influence in daily life. Health Psychol 2019 Jul;38(7):658-667 [FREE Full text] [doi: 10.1037/hea0000732] [Medline: 31008644]
5. Baym NK, Zhang YB, Lin M. Social interactions across media. New Media Soc 2016 Jun 30;6(3):299-318. [doi: 10.1177/1461444804041438]
6. Geusens F, Beullens K. The association between social networking sites and alcohol abuse among belgian adolescents. J Media Psychol 2018 Oct 30;30(4):207-216. [doi: 10.1027/1864-1105/a000196]
7. Moreno MA, D'Angelo J, Whitehill J. Social media and alcohol: summary of research, intervention ideas and future study directions. Media Comm 2016 Jun 16;4(3):50-59. [doi: 10.17645/mac.v4i3.529]
8. Hendriks H, van den Putte B, Gebhardt WA. Alcoholposts on social networking sites: the alcoholpost-typology. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):463-467. [doi: 10.1089/cyber.2017.0729] [Medline: 29995528]
9. Curtis BL, Lookatch SJ, Ramo DE, McKay JR, Feinn RS, Kranzler HR. Meta-analysis of the association of alcohol-related social media use with alcohol consumption and alcohol-related problems in adolescents and young adults. Alcohol Clin Exp Res 2018 Jun;42(6):978-986 [FREE Full text] [doi: 10.1111/acer.13642] [Medline: 29786874]
10. Moreno MA, Cox ED, Young HN, Haaland W. Underage college students' alcohol displays on Facebook and real-time alcohol behaviors. J Adolesc Health 2015 Jun;56(6):646-651 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.02.020] [Medline: 26003580]
11. Thompson CM, Romo LK. College students' drinking and posting about alcohol: forwarding a model of motivations, behaviors, and consequences. J Health Commun 2016 Jun;21(6):688-695. [doi: 10.1080/10810730.2016.1153763] [Medline: 27186824]
12. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD. Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. Addict Behav 2016 Jun;57:21-29. [doi: 10.1016/j.addbeh.2016.01.011] [Medline: 26835604]
13. Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, Kassel JD. Remember that? A comparison of real-time versus retrospective recall of smoking lapses. J Consult Clin Psychol 1997 Apr;65(2):292-300. [doi: 10.1037/0022-006x.65.2.292.a] [Medline: 9086693]
14. Sudman S, Bradburn NM. Response effects in surveys: a review and synthesis. Chicago: Aldine Publishing Co; 1974. URL: http://www.norc.org/PDFs/Publications/Response_Effects_in_Surveys_1974.pdf [accessed 2021-10-11]
15. Tucker JS, Miles JNV, D'Amico EJ. Cross-lagged associations between substance use-related media exposure and alcohol use during middle school. J Adolesc Health 2013 Oct;53(4):460-464 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.05.005] [Medline: 23770074]

16. Niland P, Lyons AC, Goodwin I, Hutton F. 'See it doesn't look pretty does it?' Young adults' airbrushed drinking practices on Facebook. Psychol Health 2014;29(8):877-895. [doi: 10.1080/08870446.2014.893345] [Medline: 24527709]
17. Hendriks H, Gebhardt WA, van den Putte B. Alcohol-related posts from young people on social networking sites: content and motivations. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):428-435. [doi: 10.1089/cyber.2016.0640] [Medline: 28650223]
18. Beullens K, Schepers A. Display of alcohol use on Facebook: a content analysis. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):497-503. [doi: 10.1089/cyber.2013.0044] [Medline: 23617225]
19. Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA. Social drinking on social media: content analysis of the social aspects of alcohol-related posts on Facebook and Instagram. J Med Internet Res 2018 Jun 22;20(6):e226 [FREE Full text] [doi: 10.2196/jmir.9355] [Medline: 29934290]
20. Cialdini R, Trost M. Social influence: social norms, conformity and compliance. In: Gilbert D, Fiske S, Lindzey G, editors. The Handbook Of Social Psychology. New York: McGraw-Hill; 1998:151-192.
21. Haines M, Spear SF. Changing the perception of the norm: a strategy to decrease binge drinking among college students. J Am Coll Health 1996 Nov;45(3):134-140. [doi: 10.1080/07448481.1996.9936873] [Medline: 8952206]
22. Lapinski MK, Rimal RN. An explication of social norms. Commun Theory 2005 May;15(2):127-147. [doi: 10.1111/j.1468-2885.2005.tb00329.x]
23. Smith SW, Alhabash S, Kanver D, Tan P, Viken G. Celebration drinking around the clock. Health Commun 2020 Oct;35(11):1307-1315. [doi: 10.1080/10410236.2019.1625007] [Medline: 31161811]
24. Berkowitz A. The social norms approach: theory, research, and annotated bibliography. 2004. URL: http://www.alanberkowitz.com/articles/social_norms.pdf [accessed 2021-10-11]
25. Bandura A, Walters R. Social Learning Theory. Englewood Cliffs: Prentice-Hall; 1977.
26. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]
27. Neighbors C, Dillard AJ, Lewis MA, Bergstrom RL, Neil TA. Normative misperceptions and temporal precedence of perceived norms and drinking. J Stud Alcohol 2006 Mar;67(2):290-299 [FREE Full text] [doi: 10.15288/jsa.2006.67.290] [Medline: 16562412]
28. Prentice D, Miller D. Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In: Advances in Experimental Social Psychology. Vol. 28. Cambridge: Academic Press; 1996:161-209.
29. Beullens K, Vandenbosch L. A conditional process analysis on the relationship between the use of social networking sites, attitudes, peer norms, and adolescents' intentions to consume alcohol. Media Psychol 2015 Oct 22;19(2):310-333. [doi: 10.1080/15213269.2015.1049275]
30. Ranney ML, Chang B, Freeman JR, Norris B, Silverberg M, Choo EK. Tweet now, see you in the ED later? Examining the association between alcohol-related tweets and emergency care visits. Acad Emerg Med 2016 Jul;23(7):831-834 [FREE Full text] [doi: 10.1111/acem.12983] [Medline: 27062454]
31. Erevik EK, Torsheim T, Andreassen CS, Vedaa Ø, Pallesen S. Disclosure and exposure of alcohol on social media and later alcohol use: a large-scale longitudinal study. Front Psychol 2017;8:1934 [FREE Full text] [doi: 10.3389/fpsyg.2017.01934] [Medline: 29163308]
32. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant 2011 Nov 19;47(4):2025-2047. [doi: 10.1007/s11135-011-9640-9]
33. Alhabash S, McAlister AR, Quilliam ET, Richards JI, Lou C. Alcohol's getting a bit more social: when alcohol marketing messages on facebook increase young adults' intentions to imbibe. Mass Commun Soc 2014 Dec 20;18(3):350-375. [doi: 10.1080/15205436.2014.945651]
34. Alhabash S, McAlister AR, Kim W, Lou C, Cunningham C, Quilliam ET, et al. Saw it on Facebook, drank it at the bar! Effects of exposure to Facebook alcohol ads on alcohol-related behaviors. J Interact Advert 2016 Mar 07;16(1):44-58. [doi: 10.1080/15252019.2016.1160330]
35. Noel JK. Using social media comments to reduce alcohol purchase intentions: an online experiment. Drug Alcohol Rev 2021 Sep;40(6):1047-1055. [doi: 10.1111/dar.13262] [Medline: 33677837]
36. Houben K, Wiers RW. Implicitly positive about alcohol? Implicit positive associations predict drinking behavior. Addict Behav 2008 Aug;33(8):979-986. [doi: 10.1016/j.addbeh.2008.03.002] [Medline: 18434034]
37. Van der Veer N, Boekee S, Hoekstra H, Peters O. Nationale Social Media Onderzoek 2018. Newcom Research and Consultancy. 2018. URL: https://www.bindinc.nl/wp-content/uploads/2018/04/Newcom-Nationale-Social-Media-Onderzoek-2018-3.pdf [accessed 2021-10-11]
38. van Hoof JJ, Bekkers J, van Vuuren M. Son, you're smoking on Facebook! College students' disclosures on social networking sites as indicators of real-life risk behaviors. Comput Human Behav 2014 May;34:249-257. [doi: 10.1016/j.chb.2014.02.008]
39. Goodrich B, Gabry J, Ali I, Brilleman S. Rstanarm: Bayesian applied regression modeling via Stan. 2018. URL: https://mc-stan.org/rstanarm/ [accessed 2021-10-11]
40. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian Data Analysis. Chapman and Hall/CRC: Boca Raton; 2013.

41. Cumming G. Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. New York: Routledge; 2012.
42. Levine TR. A defense of publishing nonsignificant (ns) results. Commun Res Rep 2013 Jul;30(3):270-274. [doi: 10.1080/08824096.2013.806261]
43. Wood PK, Sher KJ, Rutledge PC. College student alcohol consumption, day of the week, and class schedule. Alcohol Clin Exp Res 2007 Jul;31(7):1195-1207 [FREE Full text] [doi: 10.1111/j.1530-0277.2007.00402.x] [Medline: 17451400]
44. Brehm J. A Theory of Psychological Reactance. Cambridge: Academic Press; 1966.
45. Hendriks H, Wilmsen D, van Dalen W, Gebhardt WA. Picture me drinking: alcohol-related posts by Instagram influencers popular among adolescents and young adults. Front Psychol 2019;10:2991 [FREE Full text] [doi: 10.3389/fpsyg.2019.02991] [Medline: 32038379]
46. Borkenstein RF, Smith HW. The breathalyzer and its applications. Med Sci Law 2016 Jun 25;2(1):13-22. [doi: 10.1177/002580246200200103]
47. 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]
48. Van der Veer N, Boekee S, Hoekstra H, Peters O. Nationale Social Media Onderzoek 2020. Newcom Research & Consultancy. URL: https://docplayer.nl/storage/107/179454335/1633965135/RXJpoUXnldEtj_okFTNqNg/179454335.pdf [accessed 2021-10-11]
49. Quattrociocchi W, Scala A, Sunstein CR. Echo chambers on Facebook. URL: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2795110_code647786.pdf?abstractid=2795110&mirid=1 [accessed 2021-10-11]

Abbreviations

alcoholpost: alcohol-related SNS content
CI: credible interval
SNS: social networking site

Edited by R Kukafka; submitted 25.02.21; peer-reviewed by S Boyle, R Hyzer, B Kerr; comments to author 27.03.21; revised version received 26.08.21; accepted 28.09.21; published 11.11.21.

Please cite as:
Hendriks H, de Nooy W, Gebhardt WA, van den Putte B
Causal Effects of Alcohol-Related Facebook Posts on Drinking Behavior: Longitudinal Experimental Study
J Med Internet Res 2021;23(11):e28237
URL: https://www.jmir.org/2021/11/e28237
doi:10.2196/28237
PMID:34762061

Original Paper

Pragmatics to Reveal Intent in Social Media Peer Interactions: Mixed Methods Study

Tavleen Singh[1], MS; Sofia Olivares[1], BS; Trevor Cohen[2], MBChB, PhD; Nathan Cobb[3], MD; Jing Wang[4], RN, MPH, PhD; Amy Franklin[1], PhD; Sahiti Myneni[1], MSE, PhD

[1]School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, United States

[2]Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[3]Georgetown University Medical Center, Washington, DC, United States

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

Corresponding Author:
Tavleen Singh, MS
School of Biomedical Informatics
University of Texas Health Science Center
7000 Fannin Street
Houston, TX, 77030
United States
Phone: 1 713 500 3900
Email: tavleen.kaur.ranjit.singh@uth.tmc.edu

Abstract

Background: Online health communities (OHCs) have emerged as the leading venues for behavior change and health-related information seeking. The soul and success of these digital platforms lie in their ability to foster social togetherness and a sense of community by providing personalized support. However, we have a minimal understanding of how conversational posts in these settings lead to collaborative societies and ultimately result in positive health changes through social influence.

Objective: Our objective is to develop a content-specific and intent-sensitive methodological framework for analyzing peer interactions in OHCs.

Methods: We developed and applied a mixed-methods approach to understand the manifestation of expressions in peer interactions in OHCs. We applied our approach to describe online social dialogue in the context of two online communities, QuitNet (QN) and the American Diabetes Association (ADA) support community. A total of 3011 randomly selected peer interactions (n=2005 from QN, n=1006 from ADA) were analyzed. Specifically, we conducted thematic analysis to characterize communication content and linguistic expressions (speech acts) embedded within the two data sets. We also developed an empirical user persona based on their engagement levels and behavior profiles. Further, we examined the association between speech acts and communication themes across observed tiers of user engagement and self-reported behavior profiles using the chi-square test or the Fisher test.

Results: Although social support, the most prevalent communication theme in both communities, was expressed in several subtle manners, the prevalence of emotions was higher in the tobacco cessation community and assertions were higher in the diabetes self-management (DSM) community. Specific communication theme-speech act relationships were revealed, such as the social support theme was significantly associated (P<.05) with 9 speech acts from a total of 10 speech acts (ie, assertion, commissive, declarative, desire, directive, expressive, question, stance, and statement) within the QN community. Only four speech acts (ie, commissive, emotion, expressive, and stance) were significantly associated (P<.05) with the social support theme in the ADA community. The speech acts were also significantly associated with the users' abstinence status within the QN community and with the users' lifestyle status within the ADA community (P<.05).

Conclusions: Such an overlay of communication intent implicit in online peer interactions alongside content-specific theory-linked characterizations of social media discourse can inform the development of effective digital health technologies in the field of health promotion and behavior change. Our analysis revealed a rich gradient of expressions across a standardized thematic vocabulary, with a distinct variation in emotional and informational needs, depending on the behavioral and disease management profiles within and across the communities. This signifies the need and opportunities for coupling pragmatic messaging in digital therapeutics and care management pathways for personalized support.

(J Med Internet Res 2021;23(11):e32167) doi:10.2196/32167

KEYWORDS

online health communities; diabetes self-management; tobacco cessation; speech acts; behavior change; communication themes

Introduction

Lifestyle risk factors, such as tobacco use, poor diet, and physical inactivity, play an essential role in chronic disease management [1]. These health risk behaviors, albeit modifiable, result in a large number of premature deaths in the United States [2,3]. Although there are illness-specific self-management tasks, the adoption and maintenance of health behaviors are core tasks for the management of multiple chronic diseases, such as cancer, diabetes, and cardiovascular conditions [4-6]. Numerous interventions and public health campaigns have been developed to help patients incorporate new behaviors (eg, adherence to medication regimens) [7] and modify existing risky behaviors (eg, smoking cessation) [8,9] to prevent and manage chronic diseases and promote well-being. However, adherence to healthy behaviors requires significant support that targets individualistic factors and environmental influences for long time intervals [10,11].

Online health communities (OHCs) have gained popularity as venues for behavior change [12-17]. These platforms provide novel opportunities to understand complex relationships among individual actions and environmental influences in health behavior change. With the onset of mobility and connectivity in the communication sector, messages exchanged in health-related online communities reflect the intricacies of human health behavior as experienced in real time at the individual, community, and societal levels [18]. The majority of studies that have examined communication content in online communities have been limited to (1) content-agnostic analysis, such as identifying the structural characteristics of online social networks to understand how such differences might impact smoking-related behaviors [19,20]; (2) content-rich analysis, such as identifying topics, themes, and opinions from diabetes-related conversations [21,22] or capturing sentiments of individuals toward alternative smoking products [23]; and (3) content-inclusive social network analysis, such as combining content-dependent attributes with social network ties to analyze what content is being exchanged in a social tie to explain how social relationships influence behavior change [24,25]. Although several studies on health-related online communities have used manual and automated methods to analyze the content of communication, few studies have focused on the latent intent of the communication, thus ignoring the essential context relevant to health-related decision making.

In this paper, using the concepts drawn from pragmatics [26], we focus on the context of language in use, in addition to the form or content of health-related online exchanges. Pragmatic analyses enable the characterization of communication intent. Social media communication may express beliefs, ask questions, and direct another person to act. In addition to the content of the message, the context of how it is said and delivered layers further meaning to the exchange. Although pragmatics capture many aspects of semiotics (inclusive of semantics and syntax) in the transmission of a message, we narrow our scope to the speech acts prevalent in threaded discussions. Language can not only state ideas but also accomplish action through speech acts [27-29]. According to speech act theory, which was introduced by Austin [27] and later developed by Searle [28,29], speech acts capture the range of impact, or force, expressed within messages, including promises, warnings, proclamations, and statements [28,29]. Previous work on speech act analysis within online text-based communication has been applied to synchronous conversations, such as away messages on instant messengers [30], or asynchronous conversations, such as emails [31], status messages on Facebook [32,33], consumer reviews [34], forum posts [35], tweets [36,37], and emojis in WhatsApp chats [38]. Such analysis has revealed how individuals make statements, ask questions, offer suggestions, comment, and produce other speech acts that can be used to describe the strength of overall consumer sentiments through understanding implicit expressions and discourse patterns.

Although the majority of earlier works using pragmatics are in the general domain, in this paper we propose a methodological framework using speech act theory [27-29] and apply it to the messages exchanged in OHCs. Leveraging our prior work [22,39], which focused on the extraction of conversational themes and mapping of health behavior theory, we describe the relationships between theory-linked communication themes embedded in health-related peer interactions and their manifestation as individuals attempt behavior modification and self-management of chronic conditions. To ensure generalizability, our methodological framework was applied in the context of two different OHCs: (1) one for tobacco cessation and (2) another for DSM. We chose these two domains because existing research has established the influence of social relationships on risky health behaviors (eg, tobacco cessation) and lifestyle factors (eg, diet, physical activity, medication use, and self-monitoring of blood glucose) that impact type 2 diabetes mellitus (T2DM)-related health outcomes; for example, an individual is more likely to comply with health-related goals and adhere to preventive practices, provided their social ties also engage in similar behaviors by changing their intrapersonal beliefs, attitudes, or knowledge [40-43]. Through this analysis, we aim to address the following research questions:

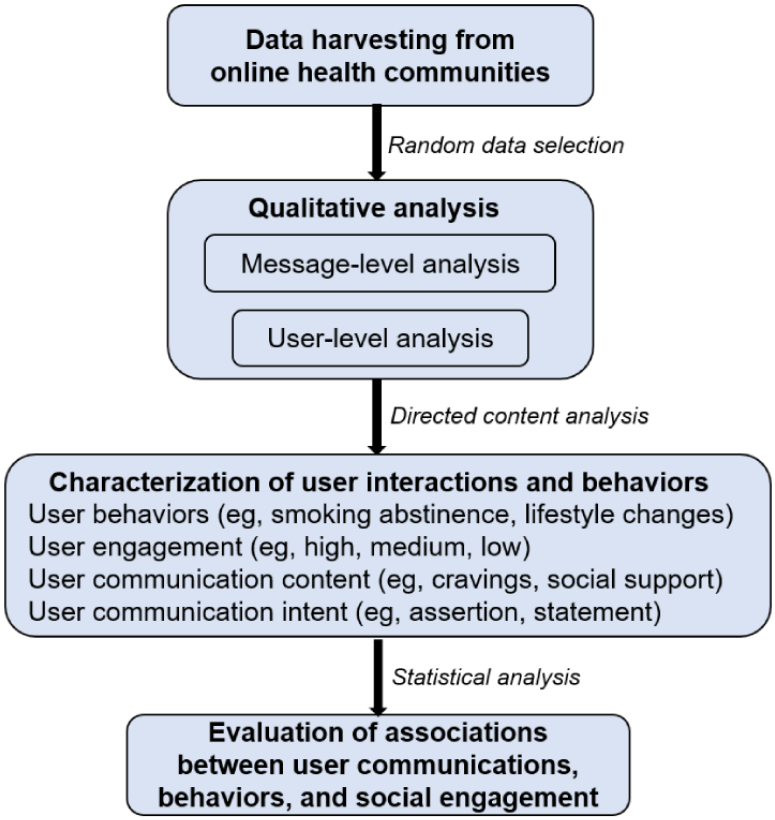
(1) How are communication content and intent expressed in OHCs for behavior change and chronic disease management?

(2) How is communication intent associated with self-reported user behavior and observed user engagement in an OHC?

Methods

Figure 1 describes the important components of our mixed-methods approach, which are explained in more detail in subsequent sections.

Figure 1. Outline of the methodological framework of the study.



Data Set Selection

The proposed research was conducted in the context of two different OHCs: one for tobacco cessation and another for DSM.

QuitNet (QN) is one of the largest OHCs promoting tobacco cessation amongst its users [44]. The users of this community are usually smokers willing to quit or ex-smokers willing to stay abstinent. Initial studies have shown a strong correlation of an individual's participation in QN with abstinence compared to individuals who do not participate in such communities [45]. The data set used in this study was drawn from a previously studied quality improvement database spanning from 2000 to 2015, consisting of nearly 2.46 million peer interactions organized into 289,456 unique threads exchanged by 64,884 unique users. Using the inbuilt sample() function of the random module in Python, we randomly selected a subset of 2005 messages from this data set, organized into 132 unique threads posted by 884 unique community users.

The *American Diabetes Association* (ADA) support community is an online support group for individuals who have diabetes (type 1, type 2, or prediabetes) to engage with their peers as well as their caregivers [46]. The community users interact with one another on a wide variety of topics ranging from medication use, diet, and physical activity to daily monitoring of blood glucose levels. Even though the outcomes among type 1 diabetes mellitus (T1DM), T2DM, or prediabetes are impacted by behaviors (eg, lifestyle, medication use, self-monitoring of blood glucose) that can be heavily influenced by an individual's social infrastructure, for this research, we focused on interactions related to T2DM. The data set spanned from 2015 to 2018, consisting of nearly 58,965 peer interactions specific to T2DM organized into 5829 unique topics posted by 1909 users. Using the inbuilt sample() function of the random module in Python, we randomly selected a subset of 1006 messages from this data set, organized into 99 unique topics posted by 141 unique community users.

Based on our prior research that focused on the qualitative analysis of digital communication from these OHCs [22,39], we expanded our data sets for this study in order to attain thematic saturation and a justifiable data set size amenable to manual analysis. Further, we selected the given size ranges for the two data sets in order to create a gold-standard annotated data set that can be efficiently used for downstream automated text mining techniques.

Qualitative Analysis

Message-Level Content and Intent Characterization

We conducted an in-depth qualitative analysis, in which we used directed content analysis [47] and discourse analysis [48] techniques, of a random sample of 3011 messages from QN and ADA online communities. Based on a modified version of Searle's taxonomy of speech acts [28,29], we classified the messages using discourse analysis into 10 categories, as described in Multimedia Appendix 1. By combining the speech acts (how) with the content (what) of these communications, we can understand how the community users accomplish their common goals, such as tobacco cessation or diabetes control—that is, how, using online exchanges, do these communities shape their identity, support their users, and accomplish the goal of helping individuals sustain positive health changes effectively and efficiently. Further, we used the coding schema provided in Multimedia Appendix 2 to map

speech acts to inductively derived, theoretically linked communication themes [22,46,49]. Communication themes capture the essence or meaning of the conversation (which essentially pertains to communication content). They are obtained through the aggregation of behavioral and cognitive constructs from multiple behavior change theories [22,46,49]. The communication theme codes were inductively derived using grounded theory techniques in our prior work [22,46]. For illustration purposes, consider the following message from the QN community to show how codes were assigned to these messages:

> *YES Cravings will go away!!! It's hard to believe, I didn't believe it either, but figured a zillion people here said they would and they were right!!! Just hang in, tough it out, stick head in freezer and breath deep and whatever you do, if any chance of slipping might creep up remember the 3 post rule!! Again, the Cravings are going to stop, it's going to be ok, I promise!!!*

This message has embodied an overarching theme of *cravings* where the user who has posted the message is providing advice on how to deal with cravings and is providing specific instructions to the other user in the form of *social support* (communication theme) as to how they can deal with this issue. The user used a *directive* speech act, where the intention was to get one's peer to remember the three-post rule, where the user is encouraged to make a post if they are craving for a smoke and wait for at least three responses. The user also used an *assertive* speech act, where the user talked about their beliefs associated with cravings, such as "Cravings are going to stop."

Communication themes and speech act codes were not mutually exclusive and may relate to multiple codes. Two independent researchers were involved in the coding process. Each coder used a qualitative coding schema to independently assign communication theme and speech act categories to a subset of messages by performing line-by-line analysis of every message using directed content analysis techniques [47]. Each message was chosen as the coding unit for analysis. Since we followed a predetermined set of codes to label the peer messages from the two communities, there was no indication for the need for new communication themes or speech acts to capture the content or intent of the communication. Interrater reliability was calculated between coders to ensure objectivity in the coding process using the Cohen kappa (κ; .78 for communication themes and .76 for speech acts) [50].

User Engagement and Behavior Characterization

We manually developed an empirical user persona based on users' engagement levels and behavior profiles. Table 1 provides an overview of the data set characteristics from both communities. The average number of messages posted by QN and ADA community users were 2 (SD=3.37) and 7 (SD=13.73), respectively. Based on this, we divided our users into three engagement levels: high-engagement users (>3 messages for QN and >=8 messages for ADA), medium-engagement users (2-3 messages for QN and 4-7 messages for ADA), and low-engagement users (1 message for QN and 1-3 messages for ADA).

Table 1. Data set characteristics.

Characteristics	QN[a]	ADA[b]
Average age of users, years (n=594)	46	—
Gender distribution (n)		
Male	130	—
Female	463	—
Not identified	291	—
User engagement status, n (%)		
High	128 (14.5)	31 (21.8)
Medium	244 (27.6)	29 (20.4)
Low	512 (57.9)	81 (57.0)
Tobacco abstinence status (n=254) specific to QN, n (%)		
Abstinent users	162 (63.8)	—
Non-abstinent users	47 (18.5)	—
Relapsed users	1	—
Users who started their quit	2	—
No self-reported status	42 (16.5)	—
DSM[c] behavior persona (n=141) specific to ADA, n (%)		
Users on medications	—	60 (42.6)
Users on no medications	—	23 (16.3)
Users with newly diagnosed T2DM[d]	—	12 (8.5)
Users with advanced T2DM	—	66 (46.8)
Users with lifestyle changes	—	36 (25.5)
Users with no lifestyle changes	—	47 (33.3)
No self-reported signatures	—	58 (41.1)

[a]QN: QuitNet.

[b]ADA: American Diabetes Association.

[c]DSM: diabetes self-management.

[d]T2DM: type 2 diabetes mellitus.

We also reported their self-reported smoking status by manually analyzing the messages, leveraging the fact that QN community users tend to specify the number of days since they last smoked in their message postings as a form of tradition. From these data, we estimated the user abstinence status for a subset of users (n=254) and identified users as falling into one of the following categories: abstinent (status 1), non-abstinent (status 0), relapsed (status 1 to status 0), users who started their quit recently (status 0 to status 1), and users with no self-reported smoking status.

Further, we constructed a DSM behavior persona for a subset of users (83/141, 58.86%) based on their self-reported forum signature and classified them as follows based on their DSM strategies and diagnostic features: (1) medication versus nonmedication users—the users were differentiated based on whether they managed their diabetes with the help of medications, such as metformin (Glucophage) or insulin, from those who did not use any medications for self-management or had no mention of medication use in their self-reported signatures; (2) users with newly diagnosed T2DM versus users with advanced T2DM—the users were differentiated based on whether they had a diagnosis of diabetes for less than 4 years (2017 or onward) from those that had had a diagnosis of diabetes for over 4 years (earlier than 2017); and (3) incorporation of lifestyle changes versus no incorporation of lifestyle changes—the users were differentiated based on whether they had incorporated lifestyle changes, such as low-carb diets or exercise, into their daily routines from those who did not incorporate such changes or had no such mentions in their self-reported signatures.

Statistical Analysis

Further, the associations among user behavior and engagement profiles with speech acts and communication themes were evaluated by statistical analysis approaches, such as the chi-square test or the Fisher test, depending on the sample size. We also used the Cramer *V* to assess their correlation strengths. All statistical analyses were performed using the R programming

language (using the stats package), and the significance level was $P<.05$.

Results

Qualitative Analysis

The qualitative analysis of the 3011 messages exchanged by QN and ADA community users provided insights into the nature of the thematic interest of the community users and the variance in the expression of intentions among them (Figures 2 and 3).

Figure 2. Prevalence of communication themes in QN and ADA communities. ADA: American Diabetes Association; QN: QuitNet.

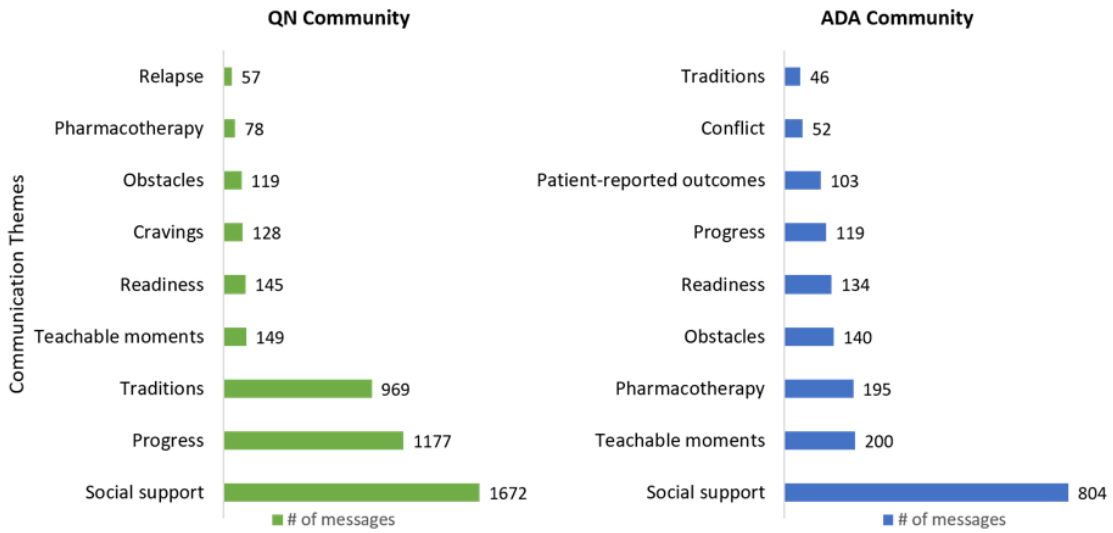
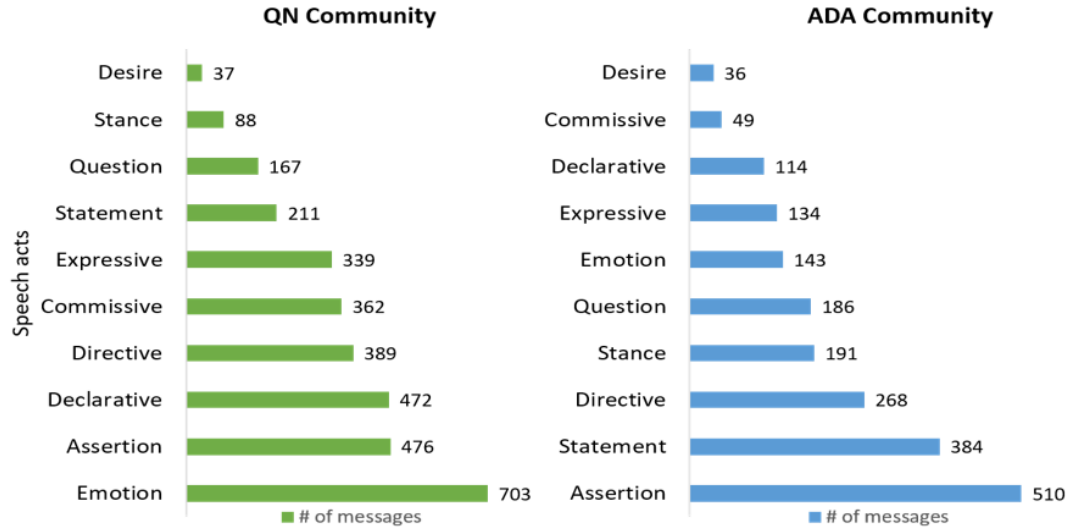


Figure 3. Prevalence of speech acts in QN and ADA communities. ADA: American Diabetes Association; QN: QuitNet.



QN Community Communication Themes

Community-driven *traditions* were quite popular in the QN community, such as bonfires, which is a virtual event in which users bring their unsmoked cigarettes and throw them into the fire, sharing stats with one another in the form of the number of days since they last smoked and celebratory exchanges upon meeting certain milestones, or pledges, which is a ritual in which users extend their hand to other community users to stay committed to abstinence. Messages describing incentives to not smoke (*teachable moments*), such as improved quality of life (eg, family bonding), better perception of everyday moments (eg, savoring food), reduction in health risks (eg, cancer), were also common in the QN community. The QN community users also posted messages seeking help to fight *cravings* or making a confession to others about their relapse or obstacles that uninspired users to stay quit. *Pharmacotherapy* options in the form of gums and patches that helped users stay away from tobacco were also commonly discussed to help those struggling in staying quit.

ADA Community Communication Themes

In the ADA community, medication-related conversations centering around the use of insulin (Lantus), metformin, etc, were quite prevalent (*pharmacotherapy*). Anxiety issues or the inability to manage blood glucose levels within the desired

range were the most commonly expressed obstacles among ADA community users (*obstacles*). The ADA community users also shared their *progress* via objective metrics, such as improved A1c values and adherence to healthy diets. *Patient-reported outcomes* were shared by ADA community users, such as higher blood glucose readings caused by beta-blockers.

QN Community Communication Intent

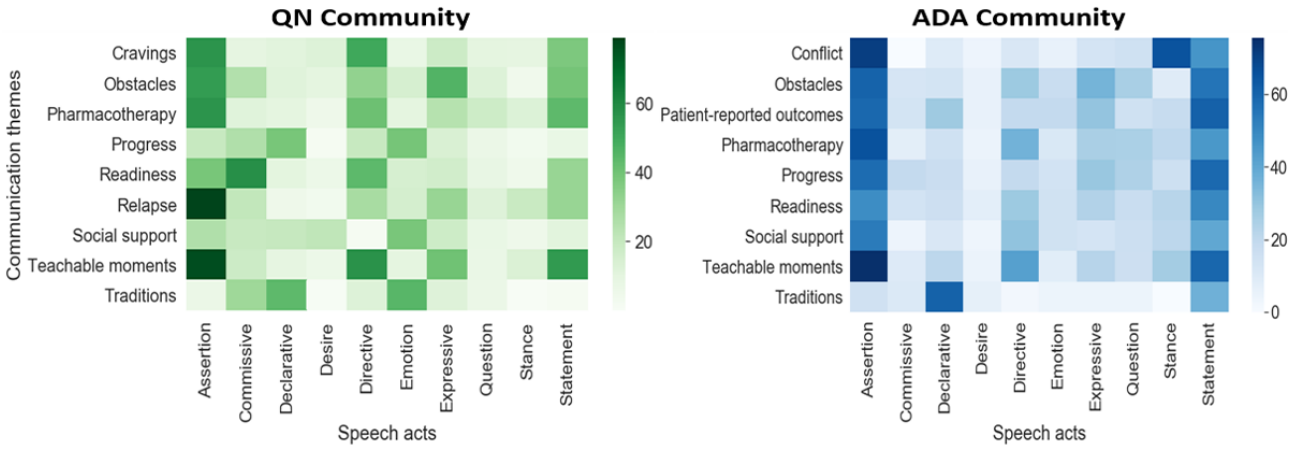
The most prevalent intention behind user communication in QN was *emotion*, which emphasizes the nature of providing emotional support to one another through such digital platforms. *Declarative* intentions where QN community users provide objective information in the form of stating their quit dates or days since the last smoke were also common. *Assertions*, such as "It really does get better, people told me that and I thought they were lying, but it really does," show how an individual's belief shapes their tobacco cessation–related goals. *Directives*, where QN community users provide support and guidance to one another, and *commissive* intentions, such as pledges or promises, were also prevalent. Some QN community users had *expressive* intentions underlying their communication with other community users, where they tried to convey their feelings about struggles with the quitting process or applauding the achievements of their peers. *Statements* such as "In my first few weeks of this quit I replaced the pack of ciggs that was always in my purse with a stack of index cards" focused on providing information about health practices that helped users stay focused on their health-related goals.

ADA Community Communication Intent

In the ADA community, *assertion* speech acts embedded within the messages, such as "Consider blurry vision as a sign of high blood sugar" or "Diet and exercise are the primary tools of defense against diabetes" were commonly present within the peer interactions. There was also a high prevalence of *statements* highlighting health-related practices of community users, such as "Since my diagnosis I have cut down carbs, started exercising and taking metformin with the goal of keeping A1C values close to normal." *Directives*, such as "Follow up with your primary care physician to get the medications checked" or "Check your blood glucose values at least before every meal in the beginning," highlighted the presence of peer support and guidance within the community. *Stance* speech acts in the form of "I agree, meds are a source of consternation" or "I disagree with your point" were also prevalent in ADA peer interactions. Many ADA community users looking for guidance from their peers posted their queries or *questions* in the forums. *Declarative* speech acts, such as "I will no longer eat or drink carelessly," where ADA community users announced objective information regarding their health-related goals, were also common. *Expressive* and *emotion* speech acts were also prevalent in ADA peer interactions to describe their efforts toward DSM or to applaud the achievements of their peers.

Figure 4 shows the co-occurrence matrix of communication themes and speech acts for both communities, where the color scale represents the percentage of messages in which a given speech act was present; the x axis represents the speech act categories, and the y axis represents the communication themes categories. In the QN community, there was a high prevalence of *commissive* speech acts in the *readiness* theme, which is expected, since this theme reflects motivations to initiate positive health changes. In addition, there was a high prevalence of *declaratives* in the *traditions* theme, since most of the traditions in the QN community focus on sharing quit statistics with other community users. In the QN community, *directives* were highly prevalent in *teachable moments* and the *cravings* theme. In the ADA community, *directives* were highly prevalent in *pharmacotherapy* theme, where users gave directions to one another on the use of medications, such as insulin and metformin. The *stance* speech act was highly prevalent in ADA messages where users' opinions were at a *conflict* with other community users, and the *expressive* speech act was commonly present in the *obstacles* theme, where the ADA community users shared their psychological state of dealing with issues such as not managing their diet plan on a day-to-day basis.

Figure 4. Mapping of communication themes and speech acts captured via peer interactions in the two OHCs. ADA: American Diabetes Association; OHC: online health community; QN: QuitNet.



Results From Statistical Analysis

The associations between all categories of speech acts and the five most prevalent categories of communication themes within the QN and ADA communities are shown in Tables 2 and 3, where *P* and V refer to the *P* value and the Cramer *V* value, respectively. In concrete terms, it means that in the QN community, the *emotion* speech act, for example, was often used in the context of the user's readiness to quit smoking or was often expressed when a user participated in community-specific *traditions* (eg, bonfires, pledges). Similarly, *stance* and *statement* speech acts, for example, were often prevalent in the discussions outlining motivations to quit smoking (*teachable moments*). *Declarative* and *directive* speech acts, for example, were often used in the context of a user's *progress* with smoking cessation. The strength of association was highest for the *desire* speech act within the *social support* theme (Cramer *V*=.26), followed by the *directive* speech act within the *social support* theme (Cramer *V*=.20) and the *declarative* speech act within the *traditions* theme (Cramer *V*=.15); see Table 2.

Table 2. Evaluation of the association between speech acts and communication themes in the QN[a] community.

	Social support, *P*(*V*)	Progress, *P*(*V*)	Traditions, *P*(*V*)	Teachable moments, *P*(*V*)	Readiness, *P*(*V*)
Assertion	<.001 (0.07)	.04 (0.02)	<.001 (0.13)	<.001 (0.11)	.02 (0.03)
Commissive	<.001 (0.07)	.71	<.001 (0.07)	<.001 (0.06)	<.001 (0.07)
Declarative	<.001 (0.11)	<.001 (0.08)	<.001 (0.15)	<.001 (0.09)	<.001 (0.08)
Desire	<.001 (0.26)	<.001 (0.13)	<.001 (0.12)	<.001 (0.04)	.02 (0.03)
Directive	<.001 (0.20)	<.001 (0.09)	.30	<.001 (0.12)	<.001 (0.10)
Emotion	.14	.73	<.001 (0.08)	<.001 (0.11)	<.001 (0.08)
Expressive	.002 (0.04)	.02 (0.03)	.02 (0.03)	<.001 (0.04)	.18
Question	.11	.51	.94	.10	.68
Stance	<.001 (0.05)	.47	<.001 (0.07)	<.001 (0.04)	.99
Statement	.05 (0.02)	.004 (0.03)	<.001 (0.11)	<.001 (0.14)	<.001 (0.07)

[a]QN: QuitNet.

In the ADA community, for example, the *assertion* speech act was mostly used in the context of a user's *readiness* to manage lifestyle behaviors that influence diabetes. The *commissive* speech act was often prevalent within the *social support*, *readiness*, and *teachable moments* communication themes. Messages conveying the user's experiences with hurdles (*teachable moments*) often contained *expressive* and *stance* speech acts. The *social support* theme often had higher prevalence of the *emotion*, *expressive*, and *stance* speech acts (see Table 3).

Table 3. Evaluation of the association between speech acts and communication themes in the ADA[a] community.

	Social support, *P*(*V*)	Progress, *P*(*V*)	Traditions, *P*(*V*)	Teachable moments, *P*(*V*)	Readiness, *P*(*V*)
Assertion	.35	.40	.74	.37	.03 (0.04)
Commissive	<.001 (0.06)	.51	.99	.04 (0.04)	.004 (0.05)
Declarative	.35	.15	.99	.49	.62
Desire	.45	.62	.99	.84	.13
Directive	.22	.51	.89	.07	.24
Emotion	.004 (0.05)	.001 (0.06)	.15	.48	.90
Expressive	<.001 (0.06)	0.74	.14	<.001 (0.07)	.29
Question	.99	.04 (0.04)	.13	.22	.67
Stance	.02 (0.04)	.63	.50	<.001 (0.07)	.84
Statement	.61	.45	.25	.45	.66

[a]ADA: American Diabetes Association.

Figure 5 shows the distribution of speech acts across the three user engagement levels for both communities, where the x axis represents the speech act categories and the y axis represents the percentage of messages in which a given speech act was present. In the QN community, low-engagement users had a high prevalence of the *desire* (2%) and *expressive* (18%) speech acts compared to high- or medium-engagement users, which indicates that low-engagement users are willing to incorporate behavior change but somehow need more motivation to engage with other users of the community who have successfully quit. *Emotion* (37%) and *declarative* (28%) speech acts were more frequently expressed by high-engagement users compared to

other engagement levels, showing how active users in the QN community play an essential role in providing emotional support to other users in this community. Statistical analysis showed that there is a significant association (P<.05) between the overall expression of speech acts and user engagement levels based on the counts of messages expressing different speech acts exchanged by individual user groups, while the overall strength of association is weak (Cramer V=.07). Post hoc analysis revealed that there is a significant association (P<.05) between the *assertion*, *declarative*, *question*, and *statement* speech acts and the user engagement levels in the QN community (see Table 4).

Figure 5. Distribution of speech acts across users' engagement levels. ADA: American Diabetes Association; QN: QuitNet.

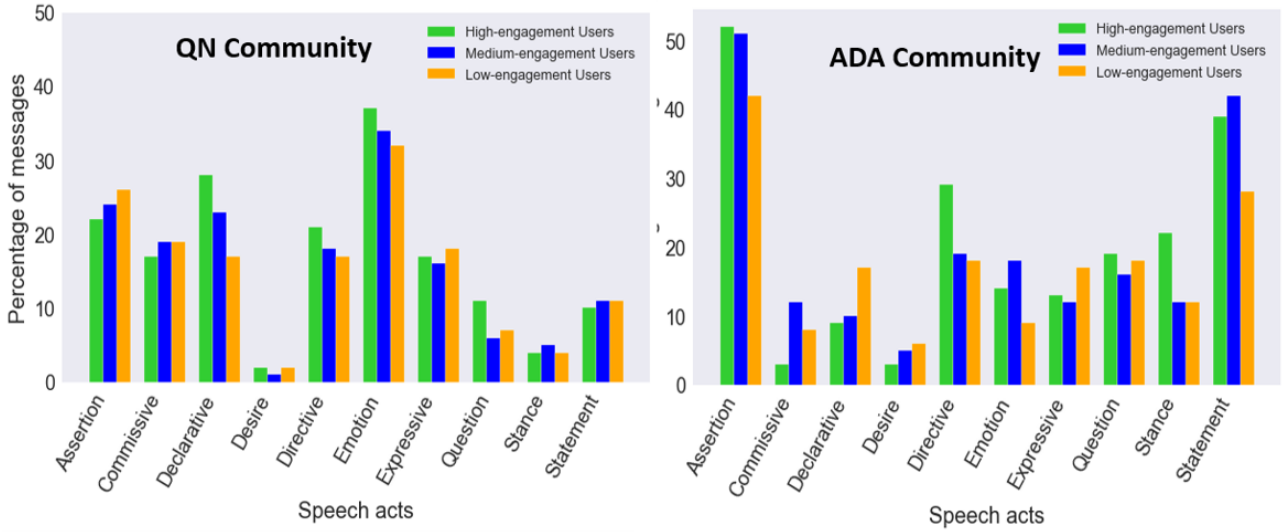


Table 4. Statistical analysis results for users' engagement levels in the QN[a] community.

Speech acts	Chi-square or Fisher exact test (P)	Cramer V
Assertion	.04	0.04
Commissive	.23	—
Declarative	.005	0.06
Desire	.34	—
Directive	.61	—
Emotion	.88	—
Expressive	.47	—
Question	.03	0.05
Stance	.77	—
Statement	<.001	0.13

[a]QN: QuitNet.

Within the ADA community, the low-engagement users expressed the *desire* (6%) and *declarative* (17%) speech acts to a greater extent compared to other engagement levels. The high-engagement users of the ADA community provided *directives* (29%) to other users compared to other engagement levels, which indicates that these users might be able to appropriately guide other community users per users' concerns. High-engagement users within the ADA community were also quite opinionated as they expressed high levels of *stance* (22%). Statistical analysis showed that there is a significant association (P<.05) between the overall expression of speech acts and user engagement levels, while the overall strength of association is weak (Cramer V=.13). Post hoc analysis revealed that there is a significant association (P<.05) between the *commissive*, *declarative*, *directive*, and *stance* speech acts and the user engagement levels in the ADA community (see Table 5).

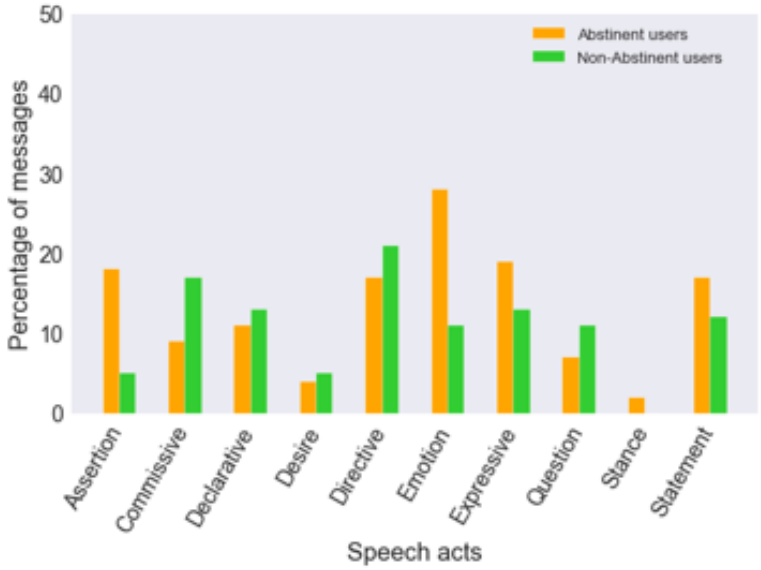
Table 5. Statistical analysis results for users' engagement levels in the ADA[a] community.

Speech acts	Chi-square or Fisher exact test (*P*)	Cramer *V*
Assertion	.80	—
Commissive	<.001	0.11
Declarative	.002	0.08
Desire	.06	—
Directive	.03	0.06
Emotion	.18	—
Expressive	.12	—
Question	.74	—
Stance	.02	0.06
Statement	.28	—

[a]ADA: American Diabetes Association.

The distribution of different speech acts by the percentage of messages exchanged per the QN community user's abstinence status is shown in Figure 6, where the x axis represents the speech act categories and the y axis represents the percentage of messages in which a given speech act was present. Abstinent users had different informational needs than non-abstinent users. Their role and context in making contributions differed from those new to the community or in a different stage of behavior change. They had a high prevalence of *emotion* (28%) and *expressive* (19%) speech acts. The other commonly occurring speech acts among abstinent users were *assertion* (18%), *directive* (17%), and *statement* (17%). The most highly prevalent speech act among non-abstinent users was *directive* (21%), and they also expressed themselves with the *commissive* (17%) and *declarative* (13%) speech acts through promises to themselves or the community. The other commonly occurring speech acts among non-abstinent users were *emotion* (13%) and *statement* (12%). The *stance* speech act was only expressed by abstinent users and had low prevalence (2%). The prevalence of the *desire* speech act was comparable across QN community users per their abstinence status—4% for abstinent users and 5% for non-abstinent users.

Figure 6. Distribution of speech acts according to the user's abstinence status in the QN community. QN: QuitNet.



We also examined the distribution of different speech acts by the percentage of messages exchanged per the ADA community users' medication status, as shown in Figure 7, where the x axis represents the speech act categories and the y axis represents the percentage of messages in which a given speech act was present. The most highly prevalent speech acts for both medication and nonmedication users were *assertion* (50% and 54%, respectively) and *statement* (both 38%). The other commonly occurring speech acts among medication and nonmedication users were *directive* (27% and 28%, respectively), *stance* (20% and 24%, respectively), and *question* (18% and 17%, respectively). Regarding the distribution of different speech acts by the percentage of messages exchanged per ADA community users' diagnosis status (Figure 7), the most highly prevalent speech acts for newly diagnosed T2DM and advanced T2DM users were *assertion* (46% and 52%, respectively) and *statement* (38% and 39%, respectively). The newly diagnosed T2DM users had a higher prevalence of the

stance (25%) speech act, while advanced T2DM users had a higher prevalence of the *directive* (29%) speech act. The users who incorporated lifestyle changes provided *directives* (29%) to others and also shared their health practices through the use of *statements* (40%). The users who incorporated lifestyle changes had a high prevalence of the *assertion* (54%), *question* (21%), and *stance* (26%) speech acts in their peer interactions compared with users who had not incorporated lifestyle changes (Figure 7).

Figure 7. Distribution of speech acts according to the user's behavior profiles in the ADA community. ADA: American Diabetes Association; T2DM: type 2 diabetes mellitus.

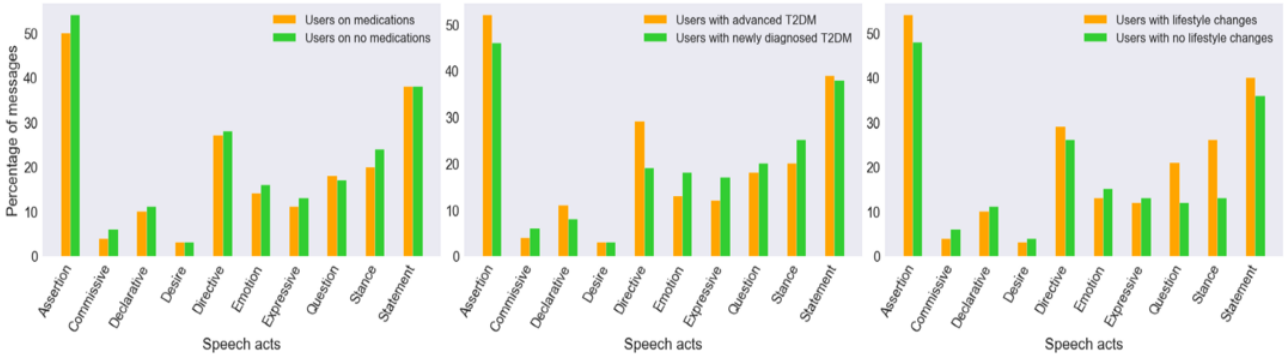


Table 6 shows the results of the statistical analysis across user behavior profiles within the two OHCs, where *P* and V refer to the *P* value and the Cramer *V* value, respectively. There was a statistically significant (*P*=.002) association between the expression of speech acts and abstinence status in the QN community. Similarly, there was a statistically significant (*P*=.03) association between the prevalence of various communication themes and abstinence status in the QN community. There was a statistically significant (*P*=.007) association between the expression of speech acts and lifestyle change status in the ADA community. This significance was also noted in the prevalence of communication themes and lifestyle change status within the ADA community (*P*=.001). The relationship strength between speech acts and abstinent status (Cramer *V*=.22) and between speech acts and lifestyle change status (Cramer *V*=.12) was weak. There was no significant relationship between the expression of speech acts and the ADA community users' medication or diagnosis status. Further analysis is required to identify the specific categories of speech acts that might contribute to these significant relationships.

Table 6. Statistical analysis results across user behavior profiles in the two OHCs[a].

	Chi-square or Fisher exact test (*P*)	Cramer *V*
Speech acts		
QN[b] (abstinent vs non-abstinent users)	.002	0.22
ADA[c] (users on medications vs users on no medications)	.95	—
ADA (users with newly diagnosed T2DM[d] vs users with advanced T2DM)	.26	—
ADA (users with lifestyle changes vs users with no lifestyle changes)	.007	0.12
Communication themes		
QN (abstinent vs non-abstinent users)	.03	0.16
ADA (users on medications vs users on no medications)	.23	—
ADA (users with newly diagnosed T2DM vs users with advanced T2DM)	.08	—
ADA (users with lifestyle changes vs users with no lifestyle changes)	<.001	0.16

[a]OHC: online health community.

[b]QN: QuitNet.

[c]ADA: American Diabetes Association.

[d]T2DM: type 2 diabetes mellitus.

Discussion

Principal Results

Our study focused on analyzing the speech acts embedded within the user communications in OHCs using the mixed-methods approach in order to gain an understanding of the factors contributing toward behavior change in the context of two health care domains, tobacco cessation and diabetes self-management (DSM). The study of communication intent to understand and model health-related user communications is not without precedent. For example, earlier research used a similar

pragmatics-based discourse analytic framework to investigate the expression of empathy within a T2DM Facebook support group and found that empathy may be expressed and perceived differently depending on the user's perspectives [51]. Similarly, in our study, although management of T2DM may include promises of behavior change through diet, exercise, and increased adherence to medication regimens, *commissive* speech acts formed only a small fraction of the discussion. Although we used the 10 categories of speech acts from Searle's taxonomy of speech acts, other studies have focused on only stance-taking intentions of complaining or disagreeing [52] or relationship-building through use of the *self-praise* speech act to understand how this speech act instills positive values among the smoking cessation community users [53]. However, it is important to look at all the categories of speech acts in order to characterize what constitutes persuasive communications so that we can better enable individuals to improve their health-related behaviors through high-impact, just-in-time adaptive support.

Another study [54] used a smoking cessation public forum called SmokingisBad to understand how information is shared between initiators/help seekers and respondents/advice givers using the analytical lens of persuasion and interpersonal pragmatics. This study reported how advice givers focus more on the interpersonal side of interaction to provide motivation and support to initiators to become successful quitters [54]. In our study, too, we noticed statistically significant differences in the expression of communication themes as well as speech acts based on the abstinence status of QN community members. The statistically significant association between the speech acts and communication themes discussed among different user behavior subgroups (abstinent vs non-abstinent users, users with lifestyle changes vs no lifestyle changes) corroborates how the intent of peer interactions and social dialogue are unique to an individual's behavior profile and also differ for a shared collection of communication themes. Abstinent users engaging with a smoking cessation community are likely to have different needs than those seeking support to stop smoking. Non-abstinent users making a promise to themselves or peers through the use of a *commissive* statement (eg, "I will stop smoking tomorrow.") logically follows. Existing tobacco cessation interventions oftentimes target an individual's readiness based on their stage of behavior change (as defined by the transtheoretical model of change [55]) and have proven to be effective in helping individuals reduce their nicotine intake [56,57]. In addition to the behavioral stage, our study enables the identification of the cognitive and behavioral state of individuals, as manifested in their peer interactions at a granularity that was not previously explored through incorporation of attributes that describe communication content and intent. The uniqueness of our study lies in understanding the intent of the community users, not just based on one's conversations with their peers, but also based on their self-reported behavior profiles as well as observed engagement levels, thus offering intrapersonal and interpersonal contexts of psychosocial and behavioral domains crucial to self-management of risky behaviors and chronic health conditions.

Significant associations between speech acts, communication themes, and user engagement within the two OHCs also suggest variances in preferred methods of expression and peer discussions, depending on the engagement levels of users. It is also interesting to note that ADA community users with low engagement had a higher prevalence of the *desire* speech act in their communication with peers, reflecting on their needs to fulfill their self-management goals. However, most ADA community users with high engagement had a high prevalence of the *directive* speech act, which may reflect their ability to provide guidance to their peers, thus emphasizing the role of peer mentors and patient expertise [58]. This provides an additional layer of social complexity that can be harnessed in the design of various digital health tools through the inclusion of advanced interactions in the form of recommendation engines that facilitate meaningful peer and content connections, conversational agents for guided training, etc. Such technological implementations can take into consideration the specific intentions associated with the engagement status of users within an OHC in order to help low-engagement users who need motivation and social support by exposing or recommending them to high-engagement users who have been able to maintain constant engagement using such platforms or to relevant topics based on their thematic interests. The age group of the community users also affects their perceptions and engagement with technology and is another important factor to consider in order to maximize the potential of digital health tools to improve consumer engagement with such platforms to facilitate behavior change [59], which is outside the scope of our analysis due to data constraints associated with our data sets.

In terms of the associations between speech acts and communication themes, all of the speech acts (except *question*) were significantly associated with the *social support* theme within the QN community. Only a few speech acts (*commissive*, *emotion*, *expressive*, and *stance*) were significantly associated with the *social support* theme in the ADA community. As social support is delivered within the full range of speech acts, this indicates robust use of many different subtle forms of transmission, emphasizing how the messages are formed matters. Areas of differentiation in how these ideas are shared (and why it differs in areas such as *progress*, *readiness* for change, and *traditions*) indicate areas where coupling pragmatic action with message content has the potential for impact. The differences observed among online communities to discuss communication themes such as *traditions*, *obstacles*, and *patient-related outcomes* also highlight the distinctive approaches necessary to address independent communities. Previous research has shown that smokers who receive comprehensive cessation counseling over the telephone using techniques that focus on cognitive, emotional, and coping processes tend to have increased acceptance of cravings to smoke [60]. Similarly, in our study, we found that in the QN community, the *directive* speech act was highly prevalent in the *cravings* theme; thus, digital behavior change interventions can benefit from incorporating such additional insights into cravings to improve message crafting and content tailoring capabilities per the user's needs.

Implications for the Design of Digital Health Interventions

Understanding the organic evolution of interaction ingredients in health-related online social conversations facilitates the synthesis of support infrastructure, including virtual chat rooms and digital assistants in health care, and depends on the comparable and adaptive emulation of naturalistic expressions in peer interactions. Such analysis can enable (1) construction of user representations and their information needs and (2) modeling of collective social resilience and human behavior patterns in collaborative endeavors in digital settings. This deeper understanding can result in downstream technologies such as digital therapeutics and virtual coaching agents with real-time naturalistic conversational capabilities. The results of our study indicate that we can use linguistic taxonomies (eg, Searle's taxonomy of speech acts) in order to capture the latent needs of OHC users based on their interactions with their peers, alongside their topics of interest (communication content), which can subsequently result in more responsive and personalized digital health interventions to support behavior change. This can have implications for designing digital care management pathways for individuals and communities based on their self-reported digital behaviors and engagement preferences, thereby enabling individuals to make better choices, and support them in sustaining long-term behavior change, thus ultimately improving their quality of life.

Limitations

Our study is not without limitations. Although QN and ADA messages were selected at random, the relatively small and uneven sample sizes may have resulted in inaccurate representations of the overall prevalence of speech acts and thematic emphasis. However, our sample of 3011 messages using qualitative research methods is appropriate for the study objectives. Even though we used empirically sound and reliable methods for coding the peer conversations, our analysis might be limited by subjective bias in the annotation process that is inherent in human coding in qualitative research [61]. Although we extracted the behavior profiles of the users who had self-reported their abstinence status within the QN forum messages and those who had created self-reported signatures within the ADA community, such extraction may be affected by incomplete or inaccurate self-report accounts and as such may not be representative of the general population. Moreover, our analysis did not consider sociodemographic and cultural factors, which can also result in differences in expressions. Our analysis was primarily focused on text-based user communications and did not consider other aspects of online user behavior within an OHC (eg, likes, dislikes, shares), which can also provide additional insights. OHCs can also be used for spreading health-related misinformation [62], which can have a negative influence on such behaviors. However, we did not consider the manifestation of misinformation as a separate theme within our data sets, which should be formalized through content-flagging mechanisms in future studies. The timeline of the data sets used in the study is not recent (2000-2015 for QN and 2015-2018 for ADA). We will attempt to obtain current data sets in future studies; however, we believe the findings still hold since the basic tenet of the social interactions remains the same. Future work should consider coding more messages in order to ensure category saturation for speech acts and communication themes so as to curate a balanced annotated data set. Our efforts to develop semiautomated methods to scale up the application of communication themes and speech act labels to large-scale data sets using deep learning models (eg, convolutional neural networks [CNNs]) [63], bidirectional encoder representations from transformers (BERT) [64]), and social influence models [65-67] are underway.

Conclusions

Digital health platforms, specifically OHCs, provide unprecedented opportunities for researchers to transform into empathetic listeners of people's endeavors, from small-scale, 1-day-at-a-time to large-scale life-consuming behavior modifications. Such listening capabilities will help us glean the needs of individuals to initiate and sustain positive health changes. This paper proposes an analytical framework to enable deep social listening capabilities to uncover the unmet needs of individuals, thus laying the foundation for next-generation technology innovation efforts. Our efforts to overlay the communication intent implicit in peer interactions alongside content-specific theory-linked characterizations of social media discourse have provided insights into topic diversity and latent interactions of users within these topics. Such content-specific and intent-sensitive methodological framework can inform the development of cognitively enabled big data analytics and machine learning models that better harvest the digital footprint of social media users.

Acknowledgments

Research reported in this publication was supported by the National Library of Medicine and the National Cancer Institute of the National Institutes of Health (award nos. 1R01LM012974-01A1 and 3R01LM012974-02S1). The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative coding schema for speech acts.
[DOCX File , 15 KB - jmir_v23i11e32167_app1.docx]

Multimedia Appendix 2
Qualitative coding schema for communication themes.
[DOCX File , 15 KB - jmir_v23i11e32167_app2.docx]

References

1. Parekh AK, Goodman RA, Gordon C, Koh HK, HHS Interagency Workgroup on Multiple Chronic Conditions. Managing multiple chronic conditions: a strategic framework for improving health outcomes and quality of life. Public Health Rep 2011;126(4):460-471 [FREE Full text] [doi: 10.1177/003335491112600403] [Medline: 21800741]
2. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA 2004;291(10):1238-1245. [doi: 10.1001/jama.291.10.1238] [Medline: 15010446]
3. National Center for Health Statistics. FastStats: Deaths and Mortality. 2021. URL: https://www.cdc.gov/nchs/fastats/deaths.htm [accessed 2021-11-01]
4. Kripalani S, Yao X, Haynes RB. Interventions to enhance medication adherence in chronic medical conditions: a systematic review. Arch Intern Med 2007;167(6):540-550. [doi: 10.1001/archinte.167.6.540] [Medline: 17389285]
5. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]
6. Gritz ER, Vidrine DJ, Fingeret MC. Smoking cessation a critical component of medical management in chronic disease populations. Am J Prev Med 2007;33(6 Suppl):S414-S422. [doi: 10.1016/j.amepre.2007.09.013] [Medline: 18021917]
7. Willey C, Redding C, Stafford J, Garfield F, Geletko S, Flanigan T, et al. Stages of change for adherence with medication regimens for chronic disease: Development and validation of a measure. Clin Ther 2000;22(7):858-871. [doi: 10.1016/s0149-2918(00)80058-2]
8. Lancaster T, Stead L, Silagy C, Sowden A. Effectiveness of interventions to help people stop smoking: findings from the Cochrane Library. BMJ 2000;321(7257):355-358 [FREE Full text] [doi: 10.1136/bmj.321.7257.355] [Medline: 10926597]
9. Lemmens V, Oenema A, Knut I, Brug J. Effectiveness of smoking cessation interventions among adults: a systematic review of reviews. Eur J Cancer Prev 2008;17(6):535-544. [doi: 10.1097/CEJ.0b013e3282f75e48] [Medline: 18941375]
10. Strecher V, Wang C, Derry H, Wildenhaus K, Johnson C. Tailored interventions for multiple risk behaviors. Health Educ Res 2002;17(5):619-626. [doi: 10.1093/her/17.5.619] [Medline: 12408206]
11. Prochaska JO, Velicer WF, Redding C, Rossi JS, Goldstein M, DePue J, et al. Stage-based expert systems to guide a population of primary care patients to quit smoking, eat healthier, prevent skin cancer, and receive regular mammograms. Prev Med 2005;41(2):406-416. [doi: 10.1016/j.ypmed.2004.09.050] [Medline: 15896835]
12. Chou W, 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]
13. Struik L, Baskerville NB. The role of Facebook in Crush the Crave, a mobile- and social media-based smoking cessation intervention: qualitative framework analysis of posts. J Med Internet Res 2014;16(7):e170 [FREE Full text] [doi: 10.2196/jmir.3189] [Medline: 25016998]
14. Cheung Y, Chan CHH, Lai CKJ, Chan WFV, Wang MP, Li HCW, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
15. Looyestyn J, Kernot J, Boshoff K, Maher C. A web-based, social networking beginners' running intervention for adults aged 18 to 50 years delivered via a Facebook group: randomized controlled trial. J Med Internet Res 2018;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
16. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. A randomized controlled trial testing a social network intervention to promote physical activity among adolescents. BMC Public Health 2018;18(1):542 [FREE Full text] [doi: 10.1186/s12889-018-5451-4] [Medline: 29685112]
17. Du Y, Paiva K, Cebula A, Kim S, Lopez K, Li C, et al. Diabetes-related topics in an online forum for caregivers of individuals living with Alzheimer disease and related dementias: qualitative inquiry. J Med Internet Res 2020;22(7):e17851 [FREE Full text] [doi: 10.2196/17851] [Medline: 32628119]
18. Marsch L. Digital health data-driven approaches to understand human behavior. Neuropsychopharmacology 2021;46(1):191-196 [FREE Full text] [doi: 10.1038/s41386-020-0761-5] [Medline: 32653896]
19. Fu L, Jacobs MA, Brookover J, Valente TW, Cobb NK, Graham AL. An exploration of the Facebook social networks of smokers and non-smokers. PLoS One 2017;12(11):e0187332 [FREE Full text] [doi: 10.1371/journal.pone.0187332] [Medline: 29095958]
20. Zhao K, Wang X, Cha S, Cohn AM, Papandonatos GD, Amato MS, et al. A multirelational social network analysis of an online health community for smoking cessation. J Med Internet Res 2016;18(8):e233 [FREE Full text] [doi: 10.2196/jmir.5985] [Medline: 27562640]
21. Karami A, Dahl AA, Turner-McGrievy G, Kharrazi H, Shaw G. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. Int J Inf Manage 2018;38(1):1-6 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2017.08.002]

22. Myneni S, Lewis B, Singh T, Paiva K, Kim SM, Cebula AV, et al. Diabetes self-management in the age of social media: large-scale analysis of peer interactions using semiautomated methods. JMIR Med Inform 2020;8(6):e18441 [FREE Full text] [doi: 10.2196/18441] [Medline: 32602843]
23. Benson R, Hu M, Chen AT, Nag S, Zhu S, Conway M. Investigating the attitudes of adolescents and young adults towards JUUL: computational study using Twitter data. JMIR Public Health Surveill 2020;6(3):e19975 [FREE Full text] [doi: 10.2196/19975] [Medline: 32876579]
24. Myneni S, Cobb NK, Cohen T. Content-specific network analysis of peer-to-peer communication in an online community for smoking cessation. AMIA Annu Symp Proc 2016;2016:934-943 [FREE Full text] [Medline: 28269890]
25. Wu N, Brazeau A, Nakhla M, Chan D, Da Costa D, Mukerji G, et al. Type 1 diabetes mellitus virtual patient network as a peer support community: protocol for social network analysis and content analysis. JMIR Res Protoc 2020;9(8):e18714 [FREE Full text] [doi: 10.2196/18714] [Medline: 32865502]
26. Thomas JA. Meaning in Interaction: An Introduction to Pragmatics. London: Routledge; 1995:240.
27. Austin JL. How to Do Things with Words. Cambridge, MA: Harvard University Press; 1975:168.
28. Searle JR. Expression and Meaning. Studies in the Theory of Speech Acts. Cambridge, England: Cambridge University Press; 1979:187.
29. Searle JR. A classification of illocutionary acts. Lang Soc 2008;5(1):1-23. [doi: 10.1017/s0047404500006837]
30. Nastri J, Peña J, Hancock JT. The construction of away messages: a speech act analysis. J Comp Mediated Comm 2006;11(4):1025-1045. [doi: 10.1111/j.1083-6101.2006.00306.x]
31. Sappelli M, Pasi G, Verberne S, de Boer M, Kraaij W. Assessing e-mail intent and tasks in e-mail messages. Inf Sci 2016;358-359:1-17 [FREE Full text] [doi: 10.1016/j.ins.2016.03.002]
32. Carr CT, Schrock DB, Dauterman P. Speech acts within Facebook status messages. J Lang Soc Psychol 2012;31(2):176-196. [doi: 10.1177/0261927x12438535]
33. Mohamad A, Rashid RA, Yunus K, Rahman SBA, Darus S, Musa R, et al. Speech acts in the Facebook status updates posted by an apostate. IJEL 2018;8(4):226 [FREE Full text] [doi: 10.5539/ijel.v8n4p226]
34. Ordenes FV, Ludwig S, De Ruyter K, Grewal D, Wetzels M. Unveiling what is written in the stars: analyzing explicit, implicit and discourse patterns of sentiment in social media. J Consum Res 2017:ucw070. [doi: 10.1093/jcr/ucw070]
35. Bhatia S, Biyani P, Mitra P. Identifying the role of individual user messages in an online discussion and its use in thread retrieval. J Assn Inf Sci Tec 2015;67(2):276-288 [FREE Full text] [doi: 10.1002/asi.23373]
36. Zhang R, Gao D, Li W. What are tweeters doing: recognizing speech acts in twitter. In: Proceedings of the 5th AAAI Conference on Analyzing Microtext: AAAIWS'11-05. AAAI Press; 2011:86-91.
37. Nemer D. Celebrities acting up: a speech act analysis in tweets of famous people. Soc Netw 2016;05(01):1-10 [FREE Full text] [doi: 10.4236/sn.2016.51001]
38. Sampietro A. Emoji and rapport management in Spanish WhatsApp chats. J Pragmat 2019;143:109-120 [FREE Full text] [doi: 10.1016/j.pragma.2019.02.009]
39. Myneni S, Cobb N, Cohen T. In pursuit of theoretical ground in behavior change support systems: analysis of peer-to-peer communication in a health-related online community. J Med Internet Res 2016;18(2):e28 [FREE Full text] [doi: 10.2196/jmir.4671] [Medline: 26839162]
40. Christakis N, Fowler J. The collective dynamics of smoking in a large social network. N Engl J Med 2008;358(21):2249-2258 [FREE Full text] [doi: 10.1056/nejmsa0706154]
41. Umberson D, Crosnoe R, Reczek C. Social relationships and health behavior across life course. Annu Rev Sociol 2010;36:139-157 [FREE Full text] [doi: 10.1146/annurev-soc-070308-120011] [Medline: 21921974]
42. Aschbrenner K, Bobak C, Schneider EJ, Naslund JA, Brunette MF, O'Malley AJ. Egocentric social networks and smoking among adults with serious mental illness. Transl Behav Med 2018;8(4):531-539 [FREE Full text] [doi: 10.1093/tbm/ibx014] [Medline: 30016519]
43. Oser TK, Oser SM, Parascando JA, Hessler-Jones D, Sciamanna CN, Sparling K, et al. Social media in the diabetes community: a novel way to assess psychosocial needs in people with diabetes and their caregivers. Curr Diab Rep 2020;20(3):10 [FREE Full text] [doi: 10.1007/s11892-020-1294-3] [Medline: 32080765]
44. Cobb N, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
45. Graham A, Cobb NK, Raymond L, Sill SM, Young J. Effectiveness of an internet-based worksite smoking cessation intervention at 12 months. J Occup Environ Med 2007;49(8):821-828 [FREE Full text] [doi: 10.1097/JOM.0b013e3180d09e6f] [Medline: 17693778]
46. American Diabetes Association. Community Overview. URL: https://community.diabetes.org/home [accessed 2021-11-02]
47. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
48. Coulthard M, Condlin C. An Introduction to Discourse Analysis. London: Routledge; 2014:228.
49. Singh T, Wang J, Myneni S. Revealing intention in health-related peer interactions: implications for optimizing patient engagement in self-health management. AMIA Annu Symp Proc 2020;2020:1120-1129 [FREE Full text] [Medline: 33936488]

50. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988:567.
51. Pounds G, Hunt D, Koteyko N. Expression of empathy in a Facebook-based diabetes support group. Discourse Context Media 2018;25:34-43 [FREE Full text] [doi: 10.1016/j.dcm.2018.01.008]
52. Epure E, Compagno D, Salinesi C, Deneckere R, Bajec M, Žitnik S. Process models of interrelated speech intentions from online health-related conversations. Artif Intell Med 2018;91:23-38 [FREE Full text] [doi: 10.1016/j.artmed.2018.06.007] [Medline: 30030089]
53. von Rohr MTR, Locher MA. The interpersonal effects of complimenting others and self-praise in online health settings. In: Complimenting Behavior and (Self-)Praise across Social Media: New contexts and new insights. Amsterdam: John Benjamins; 2020:189-211. [doi: 10.1075/pbns.313.08rud]
54. von Rohr MTR. "You will be glad you hung onto this quit": sharing information and giving support when stopping smoking online. In: Meeting Health Information Needs Outside of Healthcare: Opportunities and Challenge. Waltham, MA: Chandos/Elsevier; 2015:263-290. [doi: 10.1016/B978-0-08-100248-3.00012-3]
55. Prochaska J, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48 [FREE Full text] [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
56. Balmumcu A, Ünsal Atan Ş. Smoking cessation programs for pregnant women. UJAN 2021;32(3):188-196 [FREE Full text] [doi: 10.1097/jan.0000000000000417]
57. Ramo D, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a Facebook smoking cessation intervention for young adults. J Med Internet Res 2015;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
58. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
59. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
60. Schuck K, Otten R, Kleinjan M, Bricker JB, Engels RCME. Self-efficacy and acceptance of cravings to smoke underlie the effectiveness of quitline counseling for smoking cessation. Drug Alcohol Depend 2014;142:269-276 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.06.033] [Medline: 25042212]
61. Austin Z, Sutton J. Qualitative research: getting started. Can J Hosp Pharm 2014;67(6):436-440 [FREE Full text] [doi: 10.4212/cjhp.v67i6.1406] [Medline: 25548401]
62. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
63. Kim Y. Convolutional Neural Networks for Sentence Classification. : Association for Computational Linguistics; 2014 Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar p. 1746-1751 URL: https://aclanthology.org/D14-1181/ [doi: 10.3115/v1/D14-1181]
64. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. : Association for Computational Linguistics; 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 2019; Minneapolis, Minnesota p. 4171-4176 URL: https://aclanthology.org/N19-1423/ [doi: 10.18653/v1/N19-1423]
65. Fujimoto K, Chou C, Valente TW. The network autocorrelation model using two-mode data: affiliation exposure and potential bias in the autocorrelation parameter. Soc Netw 2011;33(3):231-243 [FREE Full text] [doi: 10.1016/j.socnet.2011.06.001] [Medline: 21909184]
66. Fujimoto K, Unger J, Valente T. A network method of measuring affiliation-based peer influence: assessing the influences of teammates' smoking on adolescent smoking. Child Dev 2012;83(2):442-451 [FREE Full text] [doi: 10.1111/j.1467-8624.2011.01729.x] [Medline: 22313152]
67. Myneni S, Fujimoto K, Cobb N, Cohen T. Content-driven analysis of an online community for smoking cessation: integration of qualitative techniques, automated text analysis, and affiliation networks. Am J Public Health 2015;105(6):1206-1212 [FREE Full text] [doi: 10.2105/ajph.2014.302464]

Abbreviations

ADA: American Diabetes Association
DSM: diabetes self-management
OHC: online health community
QN: QuitNet
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 16.07.21; peer-reviewed by C Golz, YM Huang; comments to author 09.08.21; revised version received 04.10.21; accepted 07.10.21; published 17.11.21.

Please cite as:
Singh T, Olivares S, Cohen T, Cobb N, Wang J, Franklin A, Myneni S
Pragmatics to Reveal Intent in Social Media Peer Interactions: Mixed Methods Study
J Med Internet Res 2021;23(11):e32167
URL: https://www.jmir.org/2021/11/e32167
doi:10.2196/32167
PMID:34787578

Original Paper

Noncontact Sleep Monitoring With Infrared Video Data to Estimate Sleep Apnea Severity and Distinguish Between Positional and Nonpositional Sleep Apnea: Model Development and Experimental Validation

Sina Akbarian[1,2,3], MASc; Nasim Montazeri Ghahjaverestan[1,2], PhD; Azadeh Yadollahi[1,2*], PhD; Babak Taati[1,2,3,4*], PhD

[1]Kite Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[3]Vector Institute, Toronto, ON, Canada

[4]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Babak Taati, PhD
Kite Research Institute
Toronto Rehabilitation Institute
University Health Network
550 University Ave
Toronto, ON, M5G 2A2
Canada
Phone: 1 416 597 3422 ext 7972
Email: babak.taati@uhn.ca

Abstract

Background: Sleep apnea is a respiratory disorder characterized by frequent breathing cessation during sleep. Sleep apnea severity is determined by the apnea-hypopnea index (AHI), which is the hourly rate of respiratory events. In positional sleep apnea, the AHI is higher in the supine sleeping position than it is in other sleeping positions. Positional therapy is a behavioral strategy (eg, wearing an item to encourage sleeping toward the lateral position) to treat positional apnea. The gold standard of diagnosing sleep apnea and whether or not it is positional is polysomnography; however, this test is inconvenient, expensive, and has a long waiting list.

Objective: The objective of this study was to develop and evaluate a noncontact method to estimate sleep apnea severity and to distinguish positional versus nonpositional sleep apnea.

Methods: A noncontact deep-learning algorithm was developed to analyze infrared video of sleep for estimating AHI and to distinguish patients with positional vs nonpositional sleep apnea. Specifically, a 3D convolutional neural network (CNN) architecture was used to process movements extracted by optical flow to detect respiratory events. Positional sleep apnea patients were subsequently identified by combining the AHI information provided by the 3D-CNN model with the sleeping position (supine vs lateral) detected via a previously developed CNN model.

Results: The algorithm was validated on data of 41 participants, including 26 men and 15 women with a mean age of 53 (SD 13) years, BMI of 30 (SD 7), AHI of 27 (SD 31) events/hour, and sleep duration of 5 (SD 1) hours; 20 participants had positional sleep apnea, 15 participants had nonpositional sleep apnea, and the positional status could not be discriminated for the remaining 6 participants. AHI values estimated by the 3D-CNN model correlated strongly and significantly with the gold standard (Spearman correlation coefficient 0.79, $P<.001$). Individuals with positional sleep apnea (based on an AHI threshold of 15) were identified with 83% accuracy and an F1-score of 86%.

Conclusions: This study demonstrates the possibility of using a camera-based method for developing an accessible and easy-to-use device for screening sleep apnea at home, which can be provided in the form of a tablet or smartphone app.

(J Med Internet Res 2021;23(11):e26524) doi:10.2196/26524

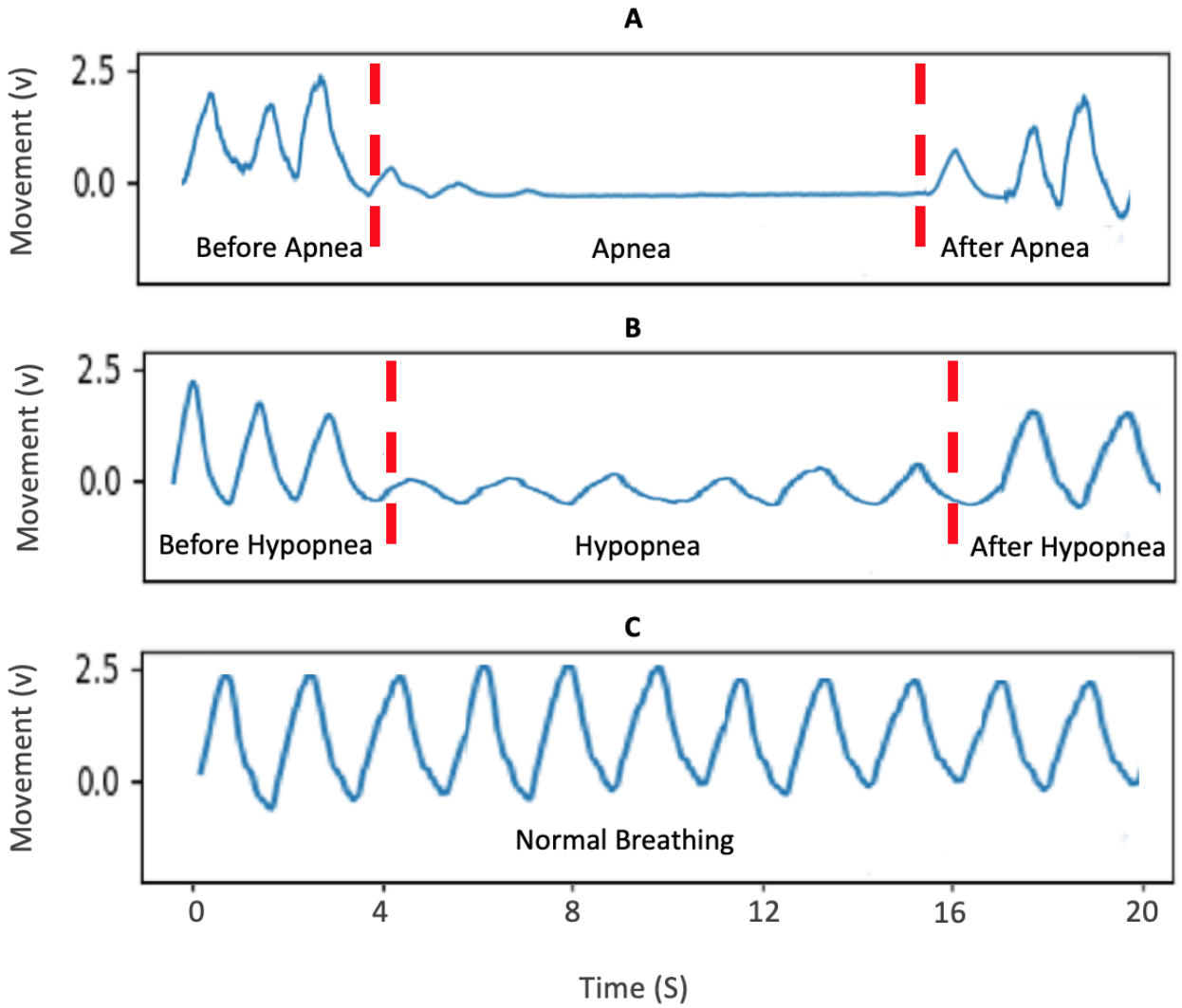
KEYWORDS

sleep apnea; deep learning; noncontact monitoring; computer vision; positional sleep apnea; 3D convolutional neural network; 3D-CNN

Introduction

Sleep apnea is a chronic respiratory disorder occurring due to frequent respiratory airflow reduction during sleep. Cessation of airflow lasting for more than 10 seconds is called apnea, whereas partial reduction in airflow by more than 30% for at least 10 seconds—in association with more than a 3% drop in blood oxygen saturation level or arousals—is called hypopnea. Sample images indicating the chest movements during normal breathing, hypopnea, and apnea are shown in Figure 1. The apnea-hypopnea index (AHI) is an indicator of the severity of sleep apnea, which measures the hourly occurrence rate of apneas and hypopneas [1]. Untreated sleep apnea raises the risk of hypertension, heart diseases, and stroke [2].

Figure 1. Sample sum of chest and abdomen movements in (A) apnea, (B) hypopnea, and (C) normal breathing.



Positional sleep apnea refers to sleep apnea patients for whom the AHI in the supine sleeping position is at least 50% higher than that in the nonsupine sleeping positions [3]. Recent studies have shown that changing to a lateral sleeping position can decrease the AHI for patients with positional sleep apnea [4]. This behavioral intervention is known as “positional therapy,” and is an effective noninvasive and nonpharmaceutical treatment for those with positional sleep apnea [5].

The current clinical approach to diagnose sleep apnea and to determine whether or not it is positional is based on polysomnography (PSG). However, PSG requires connecting more than 20 sensors to a user, which is inconvenient. A trained sleep technician manually analyzes recorded PSG signals and annotates the sleep position overnight. Moreover, PSG is expensive (>US $400) and has a long waiting time in some areas (4-36 months in Canada [6]). As a result, up to 85% of the population at risk of sleep apnea remain undiagnosed [7]. It is therefore useful to investigate screening technologies that could identify individuals at high risk via a simpler test. Increasing access to testing, diagnosis, and subsequent treatment could improve the patient’s quality of life by decreasing

hypertension and sleepiness, and can also reduce overall health care costs [8-10].

Researchers have developed several easy-to-use, convenient, and accessible methods for sleep apnea monitoring. Merchant et al [11] developed a skin-adhesive patch recording nasal pressure, blood oxygen saturation, pulse rate, respiratory effort, sleep time, and body position to estimate the AHI. Ayas et al [12] evaluated the performance of a wrist-worn device utilizing a peripheral arterial tonometer, actigraphy, and arterial oxygen saturation to diagnose sleep apnea. Varon et al [13] introduced a method for the automatic detection of sleep apnea from single-lead electrocardiogram by training a least-squares support vector machines classifier on the features extracted from the electrocardiogram signal. Several studies estimated AHI and respiratory events from analyzing tracheal sound or tracheal movements, or the combination of tracheal sound with oxygen saturation [14-18]. Lévy et al [19] utilized pulse oximetry to quantify arterial oxygen saturation and to diagnose sleep apnea.

Although these methods are more convenient than PSG, sensors attached to the body could potentially disrupt the user's regular sleep pattern. Therefore, researchers have continued to develop noncontact methods to screen individuals at risk of sleep apnea. For example, we previously developed a deep-learning model to distinguish between different types of apnea. However, as the model was not capable of detecting events, we used ground truth labels for this purpose [20]. Jakkaew et al [21] used a thermal camera to estimate breathing rate and body movements; however, they did not analyze the breathing pattern to identify sleep apnea, and the method was not designed to detect sleep position. Deng et al [22] used six active infrared cameras and a Kinect sensor to detect body position and breathing pattern (abnormal vs normal breathing). However, they did not evaluate their method in a clinical environment to demonstrate the performance for the detection of sleep apnea or positional sleep apnea. In addition, using six cameras and the Kinect will be difficult to set up in clinical or home settings, which hinders large-scale adoption. Davidovich et al [23] developed a new framework to extract the breathing pattern from a piezo-electric sensor placed under the patient's mattress through extracting time and frequency domain features and then calculating the AHI. Nandakumar et al [24] used a smartphone to emit inaudible waves and to analyze the waves' echoes from the user's body to detect respiratory events. However, these noncontact methods did not present cross-validation performance, and due to restriction in their modalities, they are not able to identify positional sleep apnea patients, which is crucial for proper treatment.

To identify patients at risk of sleep apnea and to distinguish those with positional sleep apnea, an alternative is to use computer vision and machine-learning techniques. We here propose a noncontact algorithm that analyzes infrared videos captured from a participant during sleep to estimate the AHI and to distinguish patients with positional vs nonpositional sleep apnea. Specifically, we used a 3D convolutional neural network (CNN) to analyze movements in infrared videos, to detect apneas, and to estimate the AHI. In experimental evaluation, this model outperformed a baseline model that previously reported state-of-the-art results in noncontact AHI estimation [25]. We also combined this technique with another CNN-based approach that detects the sleeping position [26] to calculate the AHI in different sleeping positions and to identify patients with positional sleep apnea. The methods and results developed in this study represent the first noncontact approach to automatically distinguish positional from nonpositional sleep apnea.

Methods

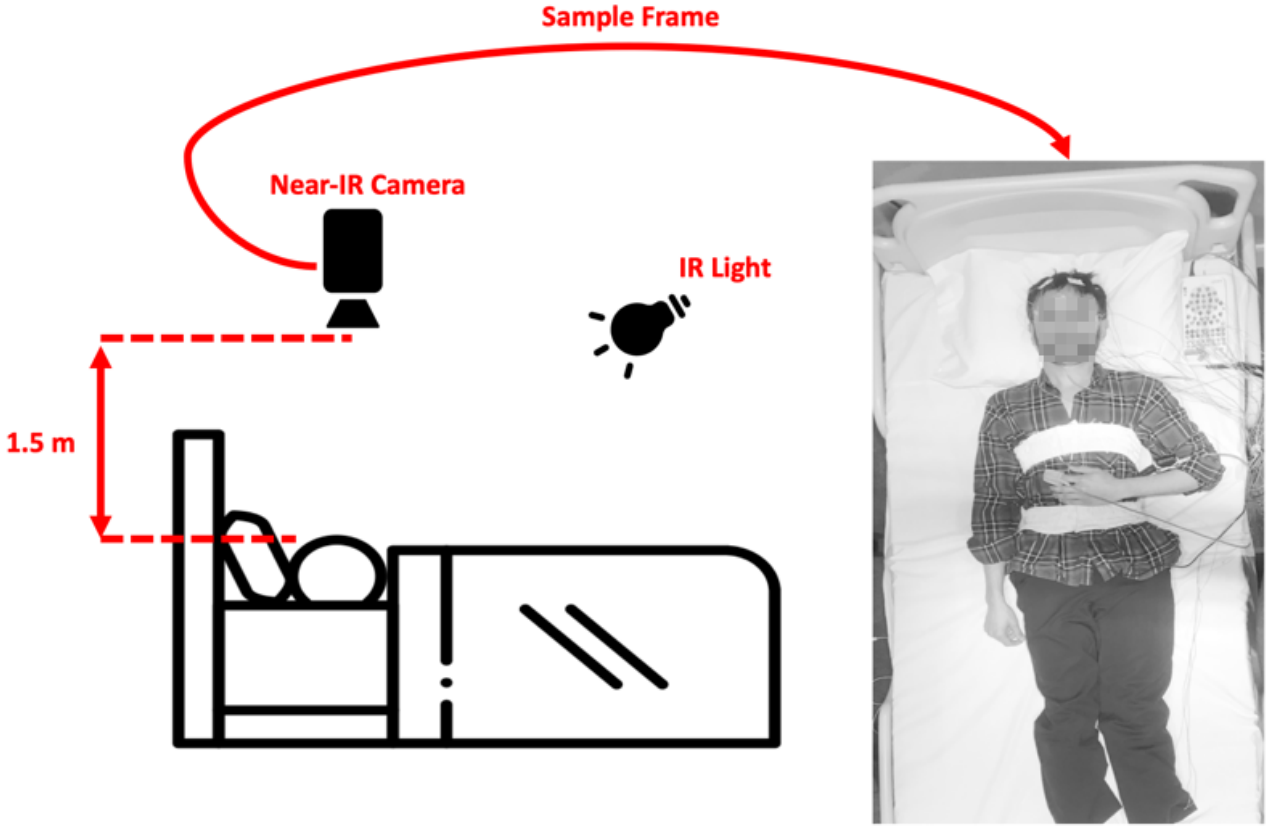
Data Collection

The University Health Network Research Ethics Board approved this study (approval number 13-7210-DE). Participants aged 18 to 85 years and without a history of cardiovascular or renal diseases were recruited for this study. Participants were recruited among patients referred for sleep diagnosis at the sleep laboratory of the Toronto Rehabilitation Institute, University Health Network. All participants signed a written consent form before taking part in the study. There were no limitations on blanket usage, movement, or clothing worn during sleep.

Simultaneously with overnight PSG (Embla s4500) that was used for a clinical diagnosis of sleep, infrared videos of participants were recorded at a resolution of 640×480 with 30 frames per second. The participants' video data were collected and synchronized with PSG signals all night for 5 (±1) hours while sleeping in a single session.

The infrared camera (Point Grey Firefly MV, 0.3 MP, FMVU-03MTM) was mounted approximately 1.5 meters above the bed. For illumination, a separate infrared light source (Raytec RM25-F-50) was mounted on the ceiling. A schematic of the camera setup and sample frame is shown in Figure 2.

Figure 2. Data collection setup and a sample anonymized image frame on the right. IR: infrared.

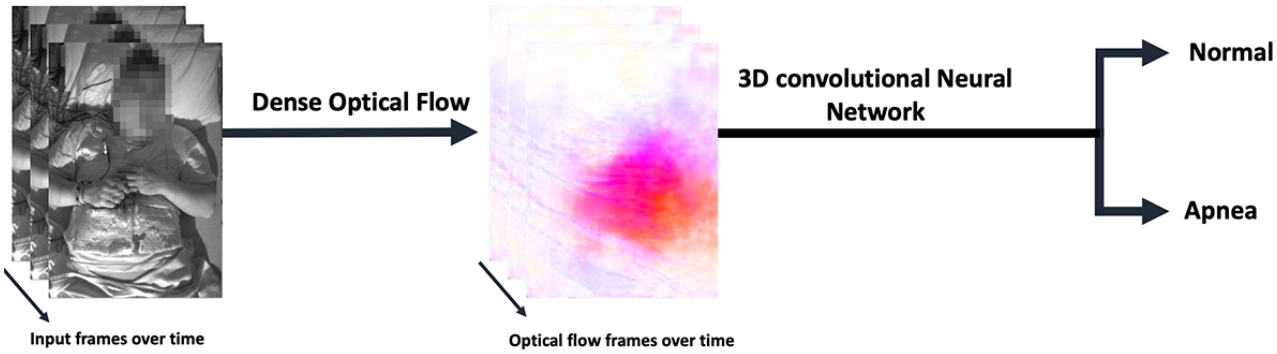


Respiratory events (apneas and hypopneas) and sleep positions (supine, lateral) of the participant throughout the night were annotated by a trained sleep technician who was blinded to the study. Since the video data were synchronized with PSG data, once the technician annotated the PSG data, all video frames were automatically labeled.

AHI Estimation

The video frames were first downsampled from 30 Hz to 2 Hz to reduce the computational cost. As breathing frequency is approximately 0.5 Hz during sleep, the reduced frequency of 2 Hz exceeds the Nyquist rate by a factor of 2. To track respiratory movements in the infrared video frames, a CNN dense optical flow (Flownet 2.0 [27]) was used, which provides accurate optical flow at a fast frame rate. Optical flow extracts movement in the *x* (side to side) and *y* (up and down) directions for each pixel in one video frame to the next. The minimum duration of an apnea is 10 seconds. This translates to 20 (or 19 in the worst case) video frames within the duration of an event. To estimate respiratory events, a 3D-CNN was trained on a sliding window of 18 optical flow images (ie, resulting from 19 consecutive video frames). Infrared videos were captured at a resolution of 640×480 pixels, resulting in optical images with a size of 640×480×2. The architecture of the 3D-CNN that was trained on the input tensors with a size of 640×480×2×18 is shown in Multimedia Appendix 1. Sample input and dense optical flow images are shown in Figure 3.

Figure 3. Sample input and dense optical flow images.



The 3D-CNN was trained with class-weighted cross-entropy loss (5 for events and 1 for normal) and the Adam optimizer. An initial value of 0.001 for the learning rate and a batch size of 25 for 25,000 epochs were chosen. The total number of parameters in this network was 8,284,265, including 8,281,829 trainable parameters and 2436 nontrainable parameters. Depending on the sleep apnea severity, respiratory events are less frequent in comparison to normal breathing; thus, the data

sets were highly imbalanced. In training time, to balance the data set, stride lengths of 0.5 and 15 seconds were used for apneas and normal breathing, respectively. In test time, a stride length of 0.5 seconds was used to predict the binary label of normal breathing versus apneas. The threshold of the trained binary classification (event vs normal) was set to 0.1 to maximize the area under the curve on the training data.

To estimate the AHI, a linear regression model was trained on the following three features: (1) the number of detected events, (2) the total duration of detected events longer than 9 seconds divided by sleep duration, and (3) sleep duration.

The performance of the 3D-CNN was compared against another approach developed by our group, which previously demonstrated state-of-the-art performance in noncontact vision-based estimation of the AHI [25]. A brief overview of this baseline approach is presented here. To extract respiratory-related motion, movements of 768 uniformly scattered points in the video frames were extracted using a sparse optical flow. Principal component analysis (PCA) was applied on the extracted point trajectories over 30-second sliding windows with a stride of 1 second to compute the predominant movements, which were associated with breathing during sleep [28]. This approach was previously validated by Zhu et al [29] and was shown to accurately track breathing rate in overnight infrared videos. To identify respiratory events from the respiratory-related motion, three features were extracted, including the respiratory rate, average power of respiratory movement, and total displacement of tracked points. Compared to normal breathing, the respiratory rate drops during respiratory events. To extract the respiratory rate, the energy of extracted respiratory movements was calculated using fast Fourier transform with a window of 10 seconds. The frequency associated with the highest energy was then considered as the respiratory rate. The second feature was the average power of respiratory movement, which decreases during a respiratory event. This feature was computed as the mean of absolute squares of respiratory displacement within a 10-second window. The last feature was total displacement, which indicates nonrespiratory movement (eg, arousals), and was determined by the summation of all of the raw optical flow movements (before applying PCA). Using these 3 features, a random forest binary classifier with 50 trees was trained to estimate sleep apnea events (apneas and hypopneas). Finally, to estimate the AHI, a linear regression model was trained using 2 features: (1) the number of predicted sleep apnea events normalized by the estimated events' duration and (2) the estimated events' duration normalized by the total sleep duration obtained from the total recording time.

Detecting Positional vs Nonpositional Sleep Apnea

For sleep position detection, a previously developed algorithm [26] was used. This method estimates body position (supine vs lateral) from a video frame using a CNN. Sample supine and lateral images are shown in Figure 4. This position detector was applied to the first video frame of each video. After each large movement (detected by thresholding the total displacement of tracked featured points extracted by optical flow over 1 second), the detector was used again to estimate the new sleeping position. As a result, a body position (supine vs lateral) was assigned to each video frame during the entire sleeping period. Once respiratory events and their associated sleep positions were detected, 6 features were calculated per person: (1) number of detected events in supine position, (2) number of detected events in lateral position, (3) total recording time in supine position, (4) total recording time in lateral position, (5) supine AHI, and (6) lateral AHI. These features were then used to train a binary random forest classifier with three trees to distinguish between positional and nonpositional sleep apnea patients.

Figure 4. Sample supine (left) and lateral (right) frames.

Validation

Leave-one-person-out cross-validation was used to evaluate the performance of AHI estimation as well as the performance of positional vs nonpositional sleep apnea detection algorithms. Bland-Altman plots and Spearman correlation coefficients were used to evaluate the performance of AHI estimation. Since an AHI of 15 is commonly used as a threshold for screening sleep apnea [30], the algorithm performance on classifying subjects as having sleep apnea was evaluated based on the threshold of AHI=15. Confusion matrices, accuracy, precision, recall, and F1-score measures were used to assess classification performance. The same measures were used to assess the performance of positional vs nonpositional sleep apnea classification.

Results

Demographic information of the 41 individuals (26 men and 15 women) recruited for this study is shown in Table 1. There were 20 participants with positional sleep apnea, 15 participants with nonpositional sleep apnea, and 6 participants that only slept in one position and as such the apnea could not be identified as either positional or nonpositional.

Table 1. Participants' demographic features for apnea-hypopnea index (AHI) estimation (N=41).[a]

Characteristics	Value, mean (SD)
Age (years)	53 (13)
BMI (kg/m^2)	30 (7)
Sleep duration (hours)	5 (1)
Number of changes in body position	9 (6)
Sleep efficiency (%)	75 (18)
REM[b] sleep percentage (%)	15 (7)
Mean wake heart rate (bpm[c])	68 (16)
Mean REM heart rate (bpm)	67 (16)
Minimum SaO_2[d]	82 (9)
Mean SaO_2	94 (3)
AHI (events/hour)	27 (31)
Supine AHI (events/hour)	41 (39)
Lateral AHI (events/hour)	21 (34)

[a]Participants' information was obtained from the sleep reports of the overnight sleep study annotated by sleep technicians.

[b]REM: rapid eye movement.

[c]bpm: beats per minute.

[d]SaO_2: arterial oxygen saturation.

The threshold used in this study for detecting position changes and ignoring the small movements (eg, breathing or pulse) was empirically set to 20,000 pixels. The total displacement was calculated by summing the displacement of all optical flow feature points [28] over 1 second and was checked against this threshold.

To evaluate the performance of AHI detection, Figure 5 and Figure 6 show the scatterplots and Bland-Altman plots between the estimated AHI and PSG-based AHI for both the 3D-CNN model and the baseline model (Zhu et al [25]).

Figure 5. Scatterplots of polysomnography (PSG) apnea-hypopnea index (AHI) vs estimated AHI values. The blue and red lines indicate fitted and unity lines, respectively. CNN: convolutional neural network.

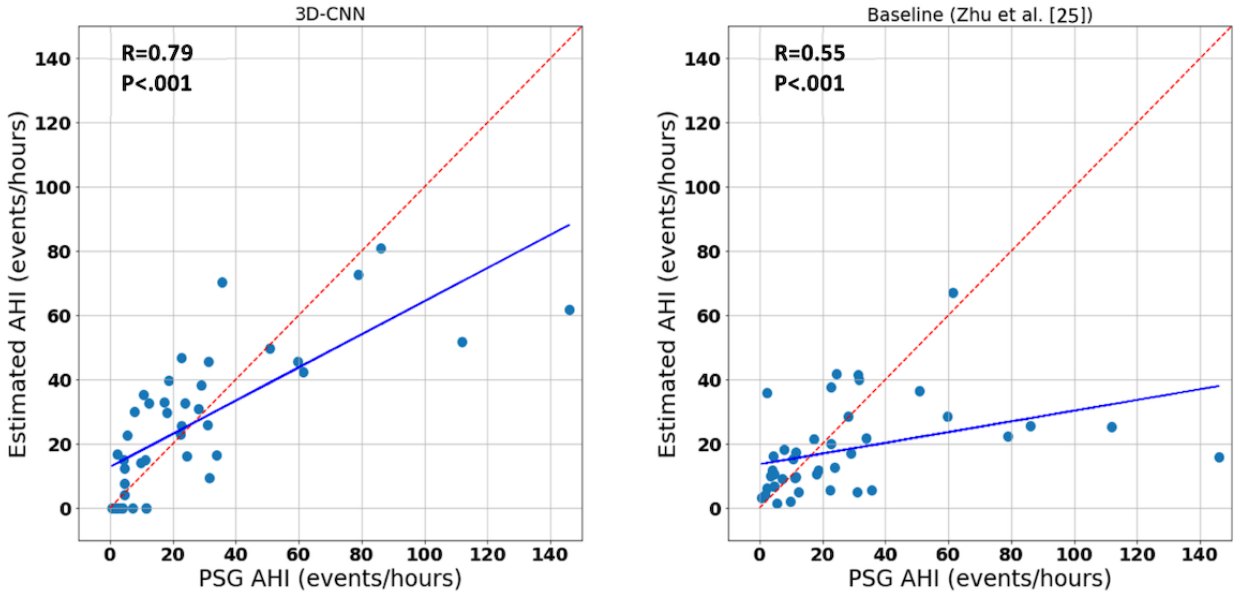
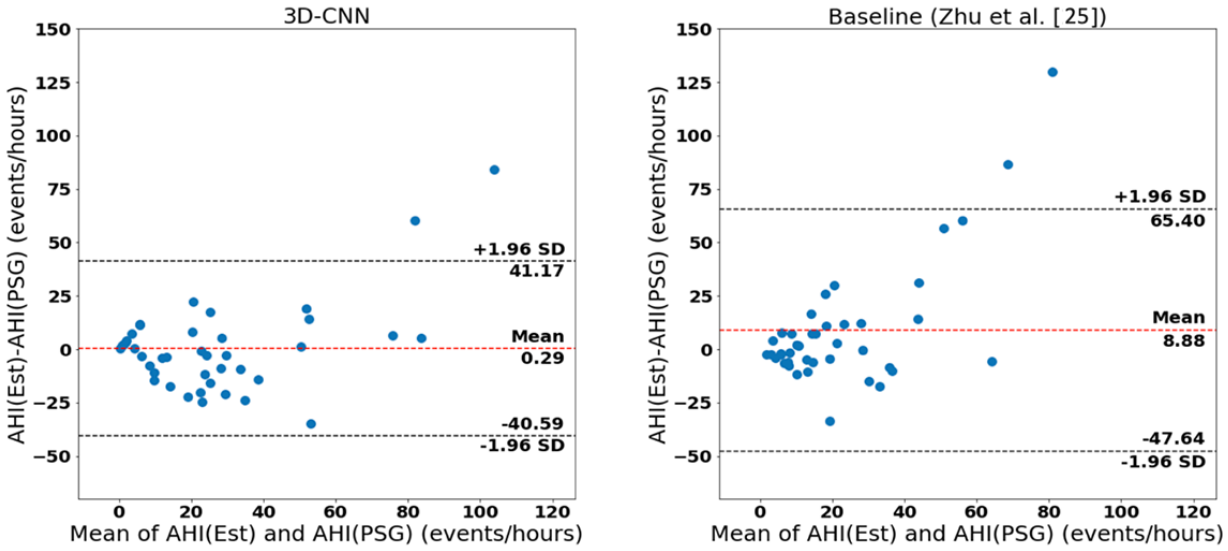


Figure 6. Bland-Altman plots of apnea-hypopnea index (AHI) estimation algorithms. PSG: polysomnography; Est: estimated; CNN: convolutional neural network.



The Spearman correlation coefficients (ρ) for AHI estimation were 0.55 and 0.79 for the baseline and 3D-CNN approach, respectively (P<.001 in both cases). In addition, the Bland-Altman plot indicated that our method outperformed the baseline according to the smaller mean (0.3 vs 8.9) and tighter 95% limits of agreement (ie, a smaller value for 1.96 of the standard deviation: 40.9 vs 56.5). Confusion matrices and the performance measures for identifying sleep apnea patients based on the AHI=15 threshold are shown in Figure 7 and in Table 2, respectively. The 3D-CNN approach obtained 83% accuracy and an F1-score of 86%, outperforming the baseline approach, which obtained an accuracy of 73% and an F1-score of 74%.

Figure 7. Confusion matrices for screening patients with sleep apnea based on the apnea-hypopnea index threshold of 15. CNN: convolutional neural network.

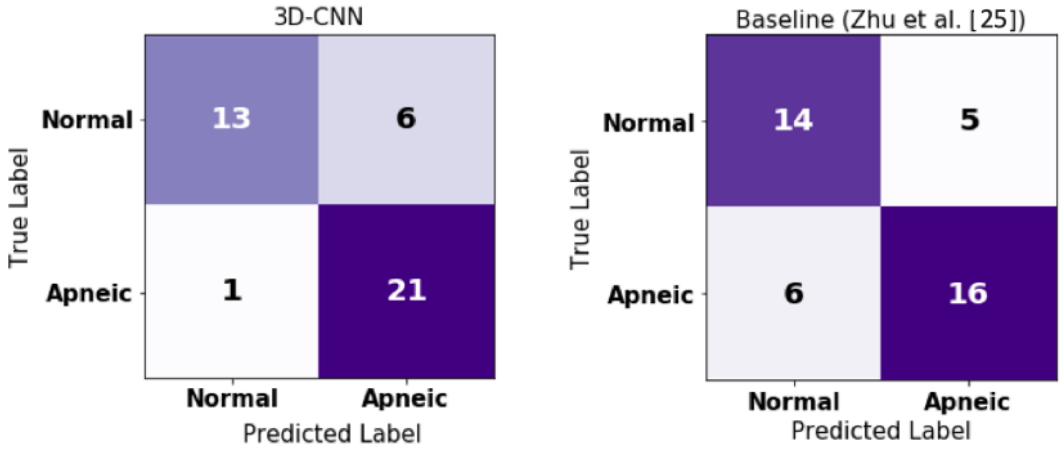


Table 2. Performance of models on screening patients with sleep apnea.

Method	Accuracy	Precision	Recall	F1-score
3D-CNN[a]	82.93	77.78	95.45	85.71
Baseline (Zhu et al [25])	73.17	76.19	72.73	74.42

[a]CNN: convolutional neural network.

The position detection algorithm estimated the body position with 83% accuracy, an F1-score of 83%, 77% precision, and 91% recall. The performance of the combination of the position detection algorithm with AHI detection on patients with positional sleep is shown in Figure 8. The 3D-CNN model classified 13 out of 20 patients with positional sleep apnea correctly. Performance measures for detecting positional vs nonpositional sleep apnea are presented in Table 3.

Figure 8. Confusion matrix for identifying positional sleep apnea. CNN: convolutional neural network.

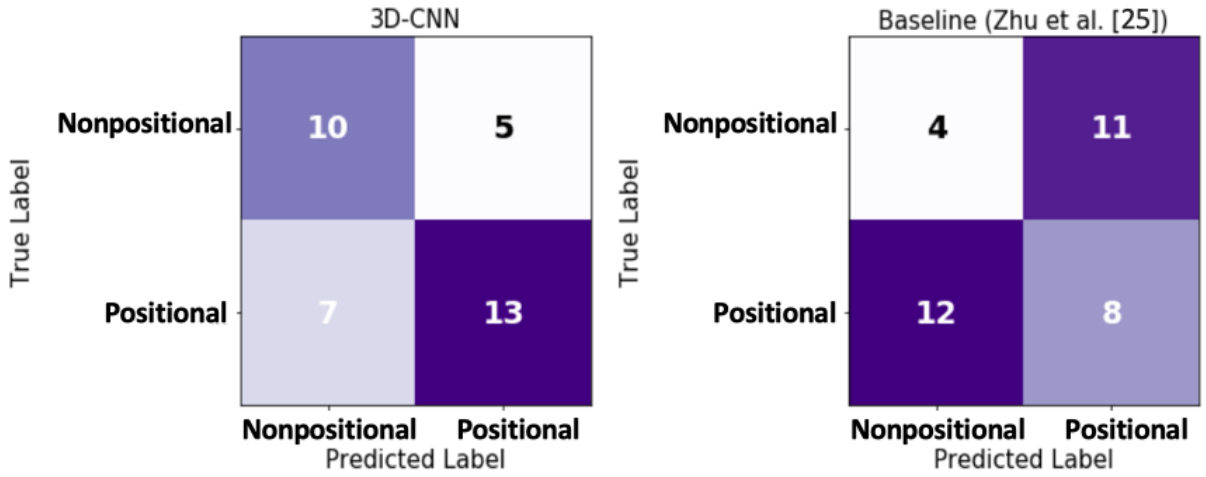


Table 3. Performance of the models in detecting positional vs nonpositional sleep apnea.

Method	Accuracy	Precision	Recall	F1-score
3D-CNN[a]	65.71	72.22	65.00	68.42
Baseline (Zhu et al [25])	34.29	42.11	40.00	41.03

[a]CNN: convolutional neural network.

Discussion

Principal Findings

The main contributions of this study are: (1) the development and experimental validation of a new noncontact approach to estimate AHI, and (2) application of this method to automatically identify individuals with positional sleep apnea. The newly developed 3D-CNN–based method outperformed the baseline model in estimating the AHI in infrared video data. However, it was ~4 times slower than the baseline algorithm. Nevertheless, the new model could still process 5 hours of sleep data in ~20 hours. Through combining estimated sleeping position information with estimated AHI, this is the first noncontact method that can identify a positional sleep apnea patient.

The developed algorithm achieved comparable performance to existing contact methods (eg, those using a single wearable sensor or a sensor placed under the mattress). For example, Hafezi et al [15] analyzed tracheal movements captured by an accelerometer to estimate AHI and to identify patients with sleep apnea. They reported a Spearman correlation of 0.86 between estimated and ground-truth (PSG) AHI values, and accuracy and F1-score values of 84% and 82%, respectively, in detecting individuals with AHI≥15. As such, they achieved a higher correlation coefficient (0.86 vs 0.79) but a lower F1-score (82% vs 86%) than our noncontact approach. An advantage of using a noncontact method over contact-based approaches is ease of use and convenience. Davidovich et al [23] used a piezo-electric sensor under a mattress to estimate the AHI. They obtained an R^2 value of 0.86 for AHI estimation, and accuracy and F1-score values of 88% and 84%, respectively, in identifying individuals with AHI≥15. Using a camera has the potential to result in a more accessible assessment technology, as it can be implemented in the form of a tablet or mobile phone app.

Limitations

Our study has some limitations. One limitation is the failure of the event detection algorithm when the participant moved out of the field of view of the camera or when the room lighting condition suddenly changed. Another limitation is the small number of participants (N=41). The algorithm was validated via leave-one-person-out cross-validation. Future work should examine the generalizability of these models to data collected in new environments.

Conclusion and Future Work

This study applied machine learning and computer vision approaches to develop a CNN-based method to detect respiratory events in different sleeping positions from data collected via an infrared camera. This method was validated on data from 41 participants to estimate AHI and to identify patients with positional sleep apnea.

This model could be used toward the development of affordable and easy-to-use technologies for screening sleep apnea at home (eg, in the form of a tablet or smartphone app). Such a system could help physicians in choosing suitable treatments for sleep apnea patients. Ultimately, improved treatment will reduce the consequences of untreated sleep apnea such as car accidents, heart disease, diabetes, and high blood pressure.

Acknowledgments

This work was supported in part by FedDev Ontario; BresoTec Inc; Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant (RGPIN-2020-04184); AMS Healthcare Fellowship in Compassion and Artificial Intelligence; and the Toronto Rehabilitation Institute, University Health Network.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Architecture of a 3D convolutional neural network used to detect apneas.
[DOCX File , 15 KB - jmir_v23i11e26524_app1.docx]

References

1. American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 1999 Aug 01;22(5):667-689. [Medline: 10450601]
2. Kim NH. Obstructive sleep apnea and abnormal glucose metabolism. Diabetes Metab J 2012 Aug;36(4):268-272 [FREE Full text] [doi: 10.4093/dmj.2012.36.4.268] [Medline: 22950056]
3. Joosten SA, O'Driscoll DM, Berger PJ, Hamilton GS. Supine position related obstructive sleep apnea in adults: pathogenesis and treatment. Sleep Med Rev 2014 Mar;18(1):7-17. [doi: 10.1016/j.smrv.2013.01.005] [Medline: 23669094]
4. Oksenberg A, Silverberg DS, Arons E, Radwan H. Positional vs nonpositional obstructive sleep apnea patients: anthropomorphic, nocturnal polysomnographic, and multiple sleep latency test data. Chest 1997 Sep;112(3):629-639. [doi: 10.1378/chest.112.3.629] [Medline: 9315794]
5. Permut I, Diaz-Abad M, Chatila W, Crocetti J, Gaughan JP, D'Alonzo GE, et al. Comparison of positional therapy to CPAP in patients with positional obstructive sleep apnea. J Clin Sleep Med 2010 Jun 15;06(03):238-243. [doi: 10.5664/jcsm.27820]
6. Flemons WW, Douglas NJ, Kuna ST, Rodenstein DO, Wheatley J. Access to diagnosis and treatment of patients with suspected sleep apnea. Am J Respir Crit Care Med 2004 Mar 15;169(6):668-672. [doi: 10.1164/rccm.200308-1124PP] [Medline: 15003950]
7. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997 Sep;20(9):705-706. [doi: 10.1093/sleep/20.9.705] [Medline: 9406321]
8. Potts KJ, Butterfield DT, Sims P, Henderson M, Shames CB. Cost savings associated with an education campaign on the diagnosis and management of sleep-disordered breathing: a retrospective, claims-based US study. Popul Health Manag 2013 Mar;16(1):7-13. [doi: 10.1089/pop.2011.0102] [Medline: 22823227]
9. Diamanti C, Manali E, Ginieri-Coccossis M, Vougas K, Cholidou K, Markozannes E, et al. Depression, physical activity, energy consumption, and quality of life in OSA patients before and after CPAP treatment. Sleep Breath 2013 Dec 6;17(4):1159-1168. [doi: 10.1007/s11325-013-0815-6] [Medline: 23386373]
10. Bratton DJ, Gaisl T, Wons AM, Kohler M. CPAP vs mandibular advancement devices and blood pressure in patients with obstructive sleep apnea: a systematic review and meta-analysis. JAMA 2015 Dec 01;314(21):2280-2293. [doi: 10.1001/jama.2015.16303] [Medline: 26624827]
11. Merchant M, Farid-Moyer M, Zobnin Y, Kuznetcov A, Savitski A, Askeland J, et al. Clinical validation of a diagnostic patch for the detection of sleep apnea. In: Sleep Medicine. 2017 Dec Presented at: SLEEP 2017: 31st Annual Meeting of the Associated Professional Sleep Societies; June 3-7, 2017; Boston, Massachusetts p. e221. [doi: 10.1016/j.sleep.2017.11.644]
12. Ayas N. Assessment of a wrist-worn device in the detection of obstructive sleep apnea. Sleep Medicine 2003 Sep;4(5):435-442. [doi: 10.1016/s1389-9457(03)00111-4]
13. Varon C, Caicedo A, Testelmans D, Buyse B, Van Huffel S. A novel algorithm for the automatic detection of sleep apnea from single-lead ECG. IEEE Trans Biomed Eng 2015 Sep;62(9):2269-2278. [doi: 10.1109/TBME.2015.2422378] [Medline: 25879836]
14. Nakano H, Hayashi M, Ohshima E, Nishikata N, Shinohara T. Validation of a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome. Sleep 2004 Aug 01;27(5):951-957. [doi: 10.1093/sleep/27.5.951] [Medline: 15453554]
15. Hafezi M, Montazeri N, Saha S, Zhu K, Gavrilovic B, Yadollahi A, et al. Sleep apnea severity estimation from tracheal movements using a deep learning model. IEEE Access 2020;8:22641-22649. [doi: 10.1109/access.2020.2969227]
16. Yadollahi A, Giannouli E, Moussavi Z. Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. Med Biol Eng Comput 2010 Nov 24;48(11):1087-1097. [doi: 10.1007/s11517-010-0674-2] [Medline: 20734154]
17. Saha S, Kabir M, Montazeri Ghahjaverestan N, Hafezi M, Gavrilovic B, Zhu K, et al. Portable diagnosis of sleep apnea with the validation of individual event detection. Sleep Med 2020 May;69:51-57. [doi: 10.1016/j.sleep.2019.12.021] [Medline: 32045854]

18. Yadollahi A, Moussavi Z. Acoustic obstructive sleep apnea detection. Annu Int Conf IEEE Eng Med Biol Soc 2009;2009:7110-7113. [doi: 10.1109/IEMBS.2009.5332870] [Medline: 19963947]
19. Lévy P, Pépin JL, Deschaux-Blanc C, Paramelle B, Brambilla C. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. Chest 1996 Mar;109(2):395-399. [doi: 10.1378/chest.109.2.395] [Medline: 8620711]
20. Akbarian S, Montazeri Ghahjaverestan N, Yadollahi A, Taati B. Distinguishing obstructive versus central apneas in infrared video of sleep using deep learning: validation study. J Med Internet Res 2020 May 22;22(5):e17252 [FREE Full text] [doi: 10.2196/17252] [Medline: 32441656]
21. Jakkaew P, Onoye T. Non-contact respiration monitoring and body movements detection for sleep using thermal imaging. Sensors (Basel) 2020 Nov 05;20(21):6307 [FREE Full text] [doi: 10.3390/s20216307] [Medline: 33167556]
22. Deng F, Dong J, Wang X, Fang Y, Liu Y, Yu Z, et al. Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor. IEEE Trans Instrum Meas 2018 Jul;67(7):1555-1563. [doi: 10.1109/tim.2017.2779358]
23. Davidovich MLY, Karasik R, Tal A, Shinar Z. Sleep apnea screening with a contact-free under-the-mattress sensor. 216 Presented at: Computing in Cardiology Conference (CinC); 2016; Vancouver, BC p. 849-852. [doi: 10.22489/CinC.2016.246-183]
24. Nandakumar R, Gollakota S, Watson N. Contactless sleep apnea detection on smartphones. GetMobile: Mobile Comp and Comm 2015 Dec 23;19(3):22-24. [doi: 10.1145/2867070.2867078]
25. Zhu K, Yadollahi A, Taati B. Non-contact apnea-hypopnea index estimation using near infrared video. In: Annu Int Conf IEEE Eng Med Biol Soc. 2019 Jul Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019; Berlin, Germany p. 792-795. [doi: 10.1109/EMBC.2019.8857711]
26. Akbarian S, Delfi G, Zhu K, Yadollahi A, Taati B. Automated non-contact detection of head and body positions during sleep. IEEE Access 2019;7:72826-72834. [doi: 10.1109/access.2019.2920025]
27. Ilg E, Mayer N, Saikia T, Dosovitskiy A, Keuper M, Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks. 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; Hawaii p. 1647-1655 URL: http://lmb.informatik.uni-freiburg.de/Publications/2017/IMSKDB17 [doi: 10.1109/CVPR.2017.179]
28. Li MH, Yadollahi A, Taati B. Noncontact vision-based cardiopulmonary monitoring in different sleeping positions. IEEE J Biomed Health Inform 2017 Sep;21(5):1367-1375. [doi: 10.1109/jbhi.2016.2567298]
29. Zhu K, Li M, Akbarian S, Hafezi M, Yadollahi A, Taati B. Vision-based heart and respiratory rate monitoring during sleep – a validation study for the population at risk of sleep apnea. IEEE J Transl Eng Health Med 2019;7:1-8. [doi: 10.1109/jtehm.2019.2946147]
30. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med 2017 Mar 15;13(3):479-504. [doi: 10.5664/jcsm.6506] [Medline: 28162150]

Abbreviations

AHI: apnea-hypopnea index
CNN: convolutional neural network
PCA: principal component analysis
PSG: polysomnography

Edited by R Kukafka, G Eysenbach; submitted 15.12.20; peer-reviewed by S Guness, W Shadid, K Pandl; comments to author 20.03.21; revised version received 13.04.21; accepted 10.09.21; published 01.11.21.

Please cite as:
Akbarian S, Ghahjaverestan NM, Yadollahi A, Taati B
Noncontact Sleep Monitoring With Infrared Video Data to Estimate Sleep Apnea Severity and Distinguish Between Positional and Nonpositional Sleep Apnea: Model Development and Experimental Validation
J Med Internet Res 2021;23(11):e26524
URL: https://www.jmir.org/2021/11/e26524
doi:10.2196/26524
PMID:34723817

Original Paper

Developing and Demonstrating the Viability and Availability of the Multilevel Implementation Strategy for Syncope Optimal Care Through Engagement (MISSION) Syncope App: Evidence-Based Clinical Decision Support Tool

Shiraz Amin[1*], BSc; Vedant Gupta[2*], MD; Gaixin Du[3], MS; Colleen McMullen[2,4], MA, MBA; Matthew Sirrine[3], BSc; Mark V Williams[5], MD, MHM; Susan S Smyth[6], MD, PhD; Romil Chadha[7], MD; Seth Stearley[8], MD; Jing Li[9], MS, MD, DrPH

[1]Performance Analytics Center of Excellence, University of Kentucky HealthCare, Lexington, KY, United States
[2]Department of Cardiovascular Medicine, University of Kentucky HealthCare, Lexington, KY, United States
[3]Center for Health Services Research, University of Kentucky, Lexington, KY, United States
[4]Gill Heart & Vascular Institute, University of Kentucky HealthCare, Lexington, KY, United States
[5]Division of Hospital Medicine, Washington University School of Medicine, St. Louis, MO, United States
[6]College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, United States
[7]Division of Hospital Medicine, University of Kentucky HealthCare, Lexington, KY, United States
[8]Department of Emergency Medicine, University of Kentucky HealthCare, Lexington, KY, United States
[9]Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States
[*]these authors contributed equally

Corresponding Author:
Jing Li, MS, MD, DrPH
Department of Medicine
Washington University School of Medicine
600 S Taylor Ave, 00155K
Campus Box 8005
St. Louis, MO, 63110
United States
Phone: 1 314 273 9386
Email: l.jing@wustl.edu

Abstract

Background: Syncope evaluation and management is associated with testing overuse and unnecessary hospitalizations. The 2017 American College of Cardiology/American Heart Association (ACC/AHA) Syncope Guideline aims to standardize clinical practice and reduce unnecessary services. The use of clinical decision support (CDS) tools offers the potential to successfully implement evidence-based clinical guidelines. However, CDS tools that provide an evidence-based differential diagnosis (DDx) of syncope at the point of care are currently lacking.

Objective: With input from diverse health systems, we developed and demonstrated the viability of a mobile app, the Multilevel Implementation Strategy for Syncope optImal care thrOugh eNgagement (MISSION) Syncope, as a CDS tool for syncope diagnosis and prognosis.

Methods: Development of the app had three main goals: (1) reliable generation of an accurate DDx, (2) incorporation of an evidence-based clinical risk tool for prognosis, and (3) user-based design and technical development. To generate a DDx that incorporated assessment recommendations, we reviewed guidelines and the literature to determine clinical assessment questions (variables) and likelihood ratios (LHRs) for each variable in predicting etiology. The creation and validation of the app diagnosis occurred through an iterative clinician review and application to actual clinical cases. The review of available risk score calculators focused on identifying an easily applied and valid evidence-based clinical risk stratification tool. The review and decision-making factors included characteristics of the original study, clinical variables, and validation studies. App design and development relied on user-centered design principles. We used observations of the emergency department workflow, storyboard demonstration, multiple mock review sessions, and beta-testing to optimize functionality and usability.

Results: The MISSION Syncope app is consistent with guideline recommendations on evidence-based practice (EBP), and its user interface (UI) reflects steps in a real-world patient evaluation: assessment, DDx, risk stratification, and recommendations. The app provides flexible clinical decision making, while emphasizing a care continuum; it generates recommendations for diagnosis and prognosis based on user input. The DDx in the app is deemed a pragmatic model that more closely aligns with real-world clinical practice and was validated using actual clinical cases. The beta-testing of the app demonstrated well-accepted functionality and usability of this syncope CDS tool.

Conclusions: The MISSION Syncope app development integrated the current literature and clinical expertise to provide an evidence-based DDx, a prognosis using a validated scoring system, and recommendations based on clinical guidelines. This app demonstrates the importance of using research literature in the development of a CDS tool and applying clinical experience to fill the gaps in available research. It is essential for a successful app to be deliberate in pursuing a practical clinical model instead of striving for a perfect mathematical model, given available published evidence. This hybrid methodology can be applied to similar CDS tool development.

(J Med Internet Res 2021;23(11):e25192) doi:10.2196/25192

KEYWORDS

cardiology; medical diagnosis; medicine; mobile applications; prognostics and health; syncope

Introduction

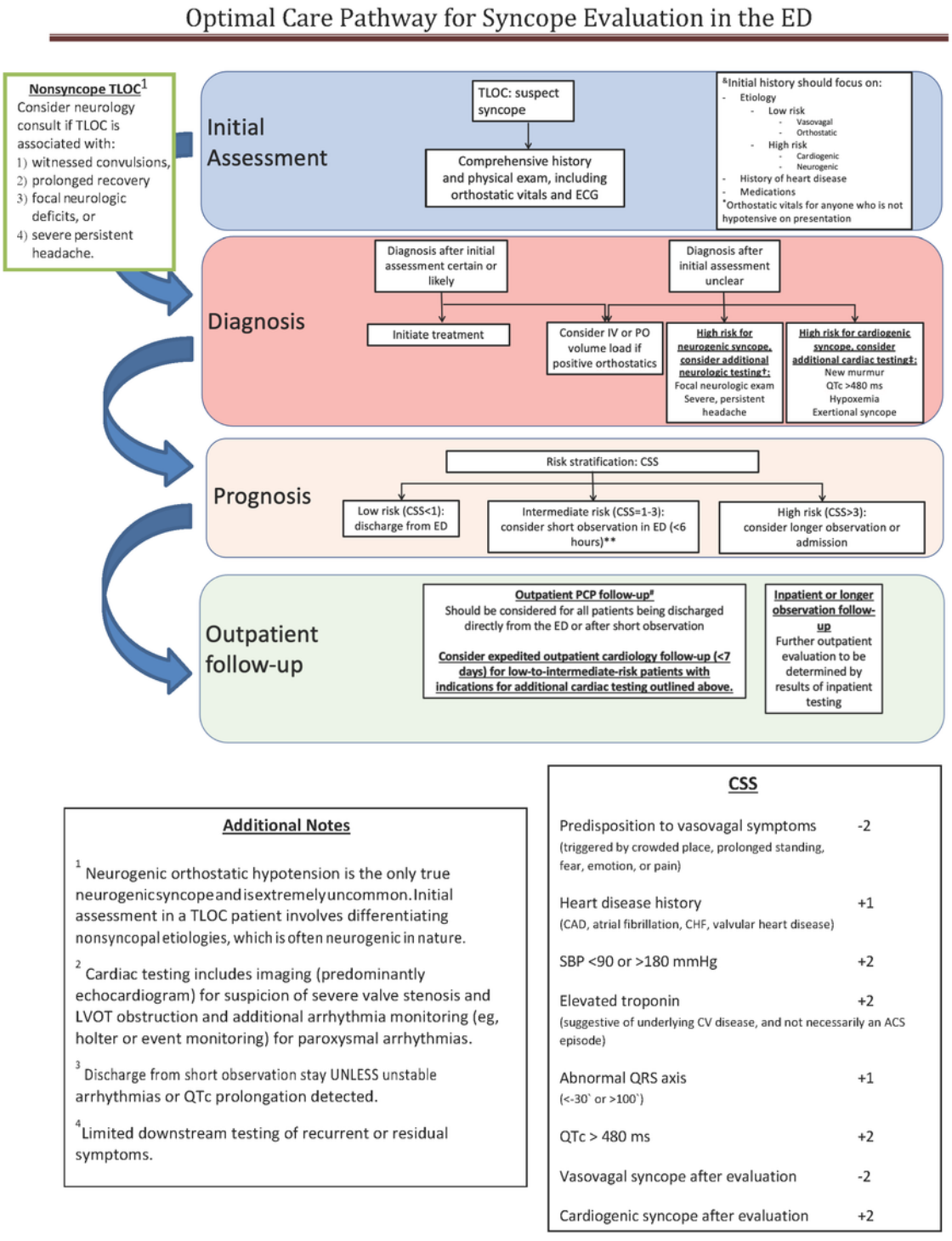
Syncope is a common yet complex presenting symptom that requires thoughtful and efficient evaluation to determine the etiology of the patient's loss of consciousness (LOC). Estimates indicate that one-half of all Americans will experience syncope during their lives, with recurrence rates as high as 13.5% [1]. The prognosis of a patient with syncope depends on the etiology and other potential underlying medical conditions. Although vasovagal reflex–mediated syncope and orthostatic hypotension are the two most common types with benign courses, a cardiac or neurologic etiology of syncope is associated with significantly higher rates of morbidity and mortality [2]. The major challenge in the evaluation of patients with syncope is that most patients are asymptomatic at the time of their presentation. Because of concerns that patients presenting with syncope are at risk for an impending catastrophic event, the overuse and inappropriate use of testing and hospital admission are common [1]. Substantial published research, including our team's work, documents the current practice of underutilization of efficient tests, overutilization of unnecessary tests, overexpenditure associated with syncope management, and heightened risk to patients due to unnecessary tests and hospitalizations [3-6]. Aiming to provide guidance on optimizing the evaluation and management of syncope, a collaboration of the American College of Emergency Physicians (ACEP), the Society for Academic Emergency Medicine (SAEM), the American College of Cardiology (ACC), the American Heart Association (AHA), and the Heart Rhythm Society (HRS) issued a *Guideline for the Evaluation and Management of Patients With Syncope* in 2017 (ie, 2017 Syncope Guideline) [7]. However, studies have found that awareness and implementation of the 2017 Syncope Guideline remain low, and current practice in the evaluation and management of syncope substantially deviates from clinical practice guidelines (CPGs) [3,8].

The use of clinical decision support (CDS) tools offers the potential to successfully implement evidence-based clinical guidelines. To assist clinicians in assessing patient risk, several syncope risk stratification calculators have been developed over the past 20 years, each with slightly different predicted clinical outcomes within various time frames [9-13]. Although risk calculators are useful, these existing tools for syncope function as medical risk calculators and do not provide any diagnosis or follow-up recommendations. CDS tools that provide an evidence-based differential diagnosis (DDx) of syncope at the point of care are currently lacking. A shared decision-making (SDM) tool for low-to-intermediate-risk patients with unexplained syncope who present to an academic emergency department (ED) in the United States is being developed but is still in the feasibility stage [14]. By incorporating CPGs and clinicians' input, CDS tools hold great potential for improving evaluation, diagnosis, care delivery, and, ultimately, outcomes in patients presenting with syncopal symptoms.

Developing a multicomponent, Multilevel Implementation Strategy for Syncope optImal care thrOugh eNgagement [MISSION]) is a multisystem implementation study aiming to adopt, adapt, and implement evidence-based practices (EBPs); engage interdisciplinary expertise; and facilitate care delivery that reduces variability, improves quality, and lowers cost. With input from diverse health systems, our study team developed the MISSION Syncope OptimalCare Pathway (Figure 1) based on the 2017 Syncope Guideline. The MISSION Syncope mobile app was designed to be a practical tool for the implementation of the MISSION Syncope OptimalCare Pathway.

The objectives of this study were to develop and demonstrate the viability of the MISSION Syncope app as a CDS tool for syncope diagnosis and prognosis that walks users through clinical assessment in a clear and concise manner consistent with EBPs and to provide recommendations based on input from the user.

Figure 1. The MISSION Syncope OptimalCare Pathway. This is the pathway to be used by clinicians to adapt and implement EBPs, engage interdisciplinary expertise, and facilitate care delivery that reduces variability, improves quality, and lowers cost. ACS: acute coronary syndrome; CAD: coronary artery disease; CHF: congestive heart failure; CSS: Canadian Syncope Score; CV, cardiovascular; EBP: evidence-based practice; ECG: electrocardiogram; ED: emergency department; IV, intravenous; LVOT: left ventricular outflow tract; MISSION: Multilevel Implementation Strategy for Syncope optImal care thrOugh eNgagement; PCP: primary care provider; PO: per os; SBP: systolic blood pressure; TLOC: transient loss of consciousness.



Methods

The development of the app was a multistep process that included (1) reliable generation of an accurate DDx, (2) incorporation of an evidence-based clinical risk tool for prognosis, and (3) user-based design and technical development. The internal review board at the University of Kentucky approved this study.

DDx

This process was guided by integrating the current literature and clinical expertise. Current guidelines recommend an initial assessment based on a comprehensive history, physical exam,

orthostatic vital signs, basic laboratory findings, and a resting electrocardiogram (ECG) [7]. These recommendations are broad and incorporate a wide range of data points and clinical indicators that complicate meaningful implementation in an app. Therefore, it was first necessary to determine relevant and proven clinical indicators that are critical in assisting diagnosis and prognosis assessment.

A literature search using PubMed and Google Scholar databases consisted of the search terms "syncope," "orthostatic hypertension," "neurogenic syncope," "vasovagal syncope," and "cardiogenic syncope." The search was limited to studies on humans. In addition, we analyzed the reference lists of identified articles to confirm we had not missed any relevant literature. The final literature search was performed on February 20, 2020. Based on the literature review, we compiled an initial list of variables; these were then reviewed by an interdisciplinary team of clinicians and consolidated into a final list. This literature review focused on the quality of study, identification of common predictors, and consolidation of study results through subject expert review and feedback. Some variables were consolidated to establish consistency and minimize different cutoffs (eg, an appropriate QT interval, ie, the time from the start of the Q wave to the end of the T wave, in different studies had varying optimal cutoff values). Focusing on app usability, the goal was to have 15-20 questions (ie, variables). This required consolidation if questions were deemed to assess a similar clinical condition and had comparable odds or likelihood ratios (LHRs). For example, a history of heart disease generically, ischemic heart disease, and atrial fibrillation were reported in separate studies with LHRs ranging from 2.4 to 7.3 [15,16]. Given the collinearity of these diseases associated with syncope, the strength of the studies, and the relative LHRs of other variables, a decision was made to consolidate these into a single question and associate the LHR from the most appropriate study.

Based on the clinical metrics identified in the literature review, we defined the inputs for a statistical model to provide an evidence-based DDx. This was done with an interdisciplinary team consisting of app developers, subject experts, and a statistician. To determine a diagnosis using a statistical model, we needed to identify the LHR for each question. This process was completed based on the quality of published research studies and subject expert (author VG) review, with special attention to variables without an LHR, a highly varying LHR, or an LHR that seemed discordant from other clinical data. This consideration included assessing the quality of original study data, sample size, patient population studied, and how the study applied to a general adult patient population.

We used the logistic regression model to calculate posttest log-odds of each cause of syncope (vasovagal, orthostatic, or cardiogenic) and included neurogenic LOC (eg, seizure) as the highest-risk nonsyncopal etiology. We defined binary logistic regression models for each etiology based on the LHRs we identified from the literature review. The use of LHRs can be beneficial in diagnosing individual patients [17]. Mathematically, a binary logistic model has a dependent variable with two possible values, which we defined as yes or no for each etiology. In the logistic model, the log-odds for the value labeled "yes" is a linear combination of one or more independent variables (Equation 1). The log-odds can then be transformed into a probability for a "yes" for each etiology (Equation 2).

For the model to work, we needed to define the β_i parameters for the model. Since we did not have sufficient data to estimate these parameter values, we instead decided to use LHRs from the existing literature. β_0 is the y intercept and is defined as the pretest odds for each etiology (Equation 3). Since this app was developed primarily for use in the ED, we decided to use the general population prevalence data for the different etiologies of syncope as our initial pretest odds for β_0. This would be mean that if we did not have any other information about the patient, the statistical model would define the posttest log-odds based on the general prevalence of each etiology. The rest of the model parameters, β_i, corresponded to each clinical indicator we had identified earlier. We then applied the LHRs (Equation 5) identified from the literature (after subject expert review and adjudication) to the appropriate questions. To simplify our model, we converted each variable into a binary (yes or no) question. Using the pretest odds ratio, the user input for each question, and the LHRs for each question, the app calculated a posttest log-odds ratio for each etiology of syncope and neurogenic LOC based on our logistic model. Based on these posttest probabilities calculated from posttest log-odds, the app subsequently displays a ranked order of each etiology as a DDx. The etiology with the highest probability is the most likely diagnosis.

For each type of syncope, a posttest probability p was calculated using logistic regression:

where

The diagnostic model was evaluated by an interdisciplinary team of experienced clinicians who assessed past cases (ie, medical chart reviews) using a web-based app developed specifically for ease of testing. The team performed a retrospective chart review of 30 patients who had presented to the ED with syncope, and compared their own diagnoses with the highest ranked differential of the statistical model. Through an iterative review and validation process, the model parameters were fine-tuned by adjusting the LHRs and the number and grouping of questions. The clinicians also identified high-risk conditions that could be missed by the diagnostic model and where a specific logic would need to be applied. The intention of this process was not to construct a perfect statistical model but rather a more pragmatic one that more closely aligns with appropriate real-world clinical practice.

The app was developed to be an adjunct to, and not a replacement for, a clinician's evaluation experience and insight; therefore, clinicians were prompted to pick their top differential (the user-selected differential) even if it did not coincide with the highest ranked app-derived differential. This created a feedback mechanism for continued improvement of the app with data on weighting of questions as well as refinement of

the mathematical formula and the parameters to the logistic regression model. Given that the Canadian Syncope Score (CSS) uses the clinician's decision in the calculation of risk stratification, the user-selected DDx, either agreeing or disagreeing with the app-generated DDx, is used for subsequent steps in the prognosis evaluation [9]. The highest ranked app-generated DDx will be used only if the user chooses not to select a diagnosis.

Incorporation of an Evidence-Based Clinical Risk Tool and Recommendation Development

The review of available risk score calculators focused on identifying the one easily applied and valid evidence-based clinical risk stratification tool. The review and decision-making factors included characteristics of the original study, clinical variables, and a validation study. After vetting other scoring systems, the Project MISSION team selected the CSS for risk stratification based on the available literature [9,18]. The CSS was chosen for its robust data and predictive capabilities; however, it is heavily focused on cardiogenic syncope [9]. This was felt to be appropriate clinically as true neurogenic syncope is rare (although neurogenic LOC is more often seen) [19]. Furthermore, the CSS has additive data on the timing of events in higher-risk patients.

The CSS uses a point system for available questions, two of which incorporate a general clinical assessment [9]. The answers to these questions were based on the user-selected differential when possible and otherwise defaulted to the app-derived differential. The remaining components were extracted from the user inputs from the assessment step, and a rapid classification into low (<1), intermediate (1-3), and high risk (>3) was conducted [9].

Next, we defined the scope of the recommendations by considering three different criteria: weighing diagnostic recommendations (what testing to do), disposition recommendations based on prognosis, and question-specific recommendations. These recommendations were generated based on the agreement or disagreement with app-derived and user-selected DDx. Recommendations were developed for all possible combinations of recommended and selected diagnoses.

User-Based Design and Technical Development

In designing the user interface (UI), we relied on user-centered design principles and specific goals were considered, including to be clear and concise, only provide information as necessary, have a logical layout, and be visually appealing. Since patients suffering syncope tend to present to the ED, we began by observing physicians in the ED and the ED workflow in general to understand the needs of users and the environment in which the app would be used. This observational information combined with more input from the practicing physicians on the team was used to create a storyboard of the patient evaluation process. This storyboard was then used to construct the various screens of the app workflow, each screen representing a different step in the process. We determined the process to involve examining the patient with questions and tests, determining a DDx, assessing the risk for adverse outcomes, and, finally, providing recommendations and performing patient follow-up.

The storyboard was worked into low-fidelity wireframes to determine the appropriate elements and display the general flow of the app. This stage of the design process also included evaluating existing medical apps that provide guidelines, protocols, and risk calculators. We evaluated the ManageAnticoag and Guideline Clinical apps, both from the ACC, and Calculate by QxMD to understand the current landscape of similar apps, determine the best practices, and decide what works and what does not [20-22]. After an iterative process, these wireframes were then worked into high-fidelity mockups using Adobe Xd (Adobe Creative Cloud; Adobe Inc., San Jose, CA, USA). Using these mockups and input from the study team, we finalized the user experience (UX) and UI for the app. The MISSION Syncope app was then developed using React Native 0.60 (Facebook Open Source; Facebook Inc., Menlo Park, CA, USA), a cross-platform mobile app development framework.

While the initial wireframes were being developed, we worked on the database and web application programming interface (API) design to be used by the app to store and retrieve data. The database was designed to store the questions, the corresponding LHRs for each kind of syncope, risk stratification, and recommendation text. In addition, the database would store each syncope evaluation performed by the user, the answer to each of the questions, the app-derived differential and user-selected differential, and the final recommendation of the app. The web API was built using C# and .NET Core with the database in Microsoft SQL Server (Microsoft Corporation, Redmond, WA, USA).

Based on input from emergency medicine (EM) and other physicians, the study team determined in the early development process that the app would not require login credentials to ensure that the clinicians would be comfortable using it and the data collected would be inherently anonymous. To maintain data integrity and secure the web API, anonymous authentication was implemented using Firebase Authentication (Google Cloud Platform; Google Inc., Menlo Park, CA, USA).

Testing was an integral part of the development life cycle. There were various phases of testing, each with a specific goal. In the first phase, we focused on testing the UI and ensuring a consistent UX across devices with varying screen resolutions, screen sizes, manufacturers, and operating systems. This was performed using the App Live testing platform (BrowserStack, Mumbai, India) on real mobile phones (iOS or Android).

In the second phase of testing, the app went through an alpha-testing phase where the app was distributed using Apple's TestFlight and Google Play's Developer testing platforms to members of the Project MISSION team. The goal of this phase was to determine usability. This allowed the team to consider the UI/UX of the app and test it over a period on their own mobile phones and provide feedback accordingly.

After making tweaks and fixes based on this testing, participants were recruited for the final beta-testing phase. These participants were primarily physicians, fellows, and residents from EM and cardiology. This phase allowed us to open up testing to a broader audience of users who were not directly involved in the development of the app, and allowed us to gather objective

feedback on usability, DDx, and performance. The beta-testers were asked to use the app for a period of several weeks, after which we conducted focus groups to obtain feedback. The focus groups were divided up based on their experiences, primarily differentiating between attending fellows or residents, and medical students. After these rounds of testing and updates to the model parameters and UI, the final version of the app was distributed through the Apple App Store and Google Play.

Results

DDx

The final list of questions, associations, and LHRs are provided in below (Tables 1 and 2). For variables that provided different cutoffs, the clinical lead identified the most appropriate one based on the literature and other sources (eg, a QT interval cutoff of 480 ms was used as it was part of the CSS) [9].

Some questions were included even if they did not have LHRs, given the importance in the DDx; were part of the CSS [9]; or were separately handled in the mathematical model. Since cardiac syncope (5%-21% of cases) and vasovagal syncope (21%-48%) are the highest causes of transient LOC (TLOC), most questions were defined for cardiogenic and vasovagal syncope [19]. The two questions included for orthostatic syncope had highly variable LHRs. One of the 2017 Guideline Class I recommendations is volume loading for patients with positive orthostatic vital signs, and given the overlap in the literature between positive orthostatic vitals and other causes of syncope, an additional question assessing the provider's clinical suspicion for an orthostatic cause of syncope was added to aid in the DDx [7]. Two questions associated with neurogenic nonsyncopal episodes were included with no LHRs because neurogenic LOC is usually not true syncope. These questions purely served as a checkpoint for routing providers to a primary neurologic workup.

The logistic regression models were found to be more complete for cardiogenic and vasovagal syncope; however, we encountered issues for neurogenic LOC and orthostatic syncope due to limited studies on a DDx for these etiologies. We derived this conclusion by performing an analysis of the data collected from app usage during the alpha- and beta-testing phases. We compared the highest-ranked DDx generated by our mathematical formula with users' selections to determine congruence with our formulas. We found that when the user determined the DDx to be cardiogenic or vasovagal, our mathematical formula derived the same result in 70% of cases. A specific logic was applied to the formula based on clinical experience. The decision was made to handle the two low-performing etiologies, orthostatic and neurogenic LOC, by applying artificial weighting to move these etiologies to the top of the differential if all questions for each type were answered with yes and by providing targeted question-specific recommendations. For example, the MISSION Syncope app recommends volume loading when a patient is orthostatic, but the app-derived or user-selected differential can be cardiogenic or vasovagal, respectively [7]. For neurogenic LOC, weighting was purely based on clinical experience, with no reference to a discrete LHR, because no consistent LHR exists in the literature that would mathematically rank this etiology to the top of the differential. In addition, since the neurogenic risk to the patient (prognosis) was not addressed by the CSS, a deliberate mention of concerning features and additive recommendations were deemed important [9].

Table 1. Final assessment questions, with LHRs[a] for vasovagal syncope.

Question	Reference	LHR+	LHR–
Is the patient less than or equal to 35 years of age?	[16]	7.29	0.30
Does the patient have a history of heart disease (atrial fibrillation/flutter, ventricular tachycardia, heart block, heart failure, stable ischemic heart disease, valvular heart disease)?[b]	[15,16,23]	0.072	1.82
Did the syncopal episode occur in the context of any of the following: warm or crowded place, prolonged standing, fear, emotion, pain, or using the toilet?[b]	[16]	8.85	0.498
Was the syncopal episode associated with chest pain?	[23,24]	NULL[c]	NULL[c]
Was the syncopal episode associated with palpitations?	[13,15,16,23-25]	NULL[c]	NULL[c]
Was the syncopal episode associated with exertion?	[13,15]	NULL[c]	NULL[c]
Was the syncopal episode associated with position change?[d]	—	NULL[c]	NULL[c]
Was the syncopal episode associated with hypoxia?	[16]	0.104	1.08
Was the syncopal episode associated with nausea, vomiting, or a warm/flushed feeling?	[13,15,16,23-25]	5.10	0.552
Does the patient describe any of the following: severe headache, focal neurologic deficit, or postictal state?[e]	[16]	NULL[c]	NULL[c]
Were there convulsions witnessed associated with the syncope?[d]	—	NULL[c]	NULL[c]
Is there a new murmur on exam?[d]	—	NULL[c]	NULL[c]
Is the resting SBP[f] <90 mmHg or >180 mmHg?	[16]	NULL[c]	NULL[c]
Were orthostatic vitals positive (>20 mmHg drop in SBP or >30 beats per minute increase in heart rate)?[g]	—	NULL[c]	NULL[c]
Do you think orthostasis is the cause for syncope?[g]	—	NULL[c]	NULL[c]
Were there any new focal neurologic deficits on physical exam?[e]	—	NULL[c]	NULL[c]
Is the QRS axis abnormal (<–30 degrees or >100 degrees)?[b]	—	NULL[c]	NULL[c]
Is the QRS duration prolonged (>120 ms)?[b]	—	NULL[c]	NULL[c]
Is the corrected QT interval prolonged (>480 ms)?[b]	—	NULL[c]	NULL[c]
Is the troponin elevated (high-sensitivity cardiac troponin *t*>14 ng/L)?[b]	[26]	NULL[c]	NULL[c]

[a]LHR: likelihood ratio.

[b]Input for the Canadian Syncope Score (CSS).

[c]Ratios not found in the literature.

[d]Included to prompt additional considerations.

[e]Artificially weighted for neurogenic loss of consciousness (LOC).

[f]SBP: systolic blood pressure.

[g]Artificially weighted for orthostatic syncope.

Table 2. Final assessment questions, with LHRs[a] for cardiogenic syncope.

Question	Reference	LHR+	LHR–
Is the patient less than or equal to 35 years of age?	[16]	0.13	3.24
Does the patient have a history of heart disease (atrial fibrillation/flutter, ventricular tachycardia, heart block, heart failure, stable ischemic heart disease, valvular heart disease)?[b]	[15,16,23]	2.93	0.74
Did the syncopal episode occur in the context of any of the following: warm or crowded place, prolonged standing, fear, emotion, pain, or using the toilet?[b]	[16]	0.167	1.43
Was the syncopal episode associated with chest pain?	[23,24]	4.25	0.881
Was the syncopal episode associated with palpitations?	[13,15,16,23-25]	3.78	0.853
Was the syncopal episode associated with exertion?	[13,15]	4.36	0.896
Was the syncopal episode associated with position change?[c]	—	NA[d]	NA
Was the syncopal episode associated with hypoxia?	[16]	3.74	0.94
Was the syncopal episode associated with nausea, vomiting, or a warm/flushed feeling?	[13,15,16,23-25]	0.354	1.38
Does the patient describe any of the following: severe headache, focal neurologic deficit, or postictal state?[e]	[16]	0.170	1.21
Were there convulsions witnessed associated with the syncope?[c]	—	NULL[f]	NULL[f]
Is there a new murmur on exam?[c]	—	NULL[f]	NULL[f]
Is the resting SBP[g] <90 mmHg or >180 mmHg?	[16]	5.88	0.894
Were orthostatic vitals positive (>20 mmHg drop in SBP or >30 beats per minute increase in heart rate)?[h]	—	NULL[f]	NULL[f]
Do you think orthostasis is the cause for syncope?[h]	—	NULL[f]	NULL[f]
Were there any new focal neurologic deficits on physical exam?[e]	—	NULL[f]	NULL[f]
Is the QRS axis abnormal (<–30 degrees or >100 degrees)?[b]	—	NULL[f]	NULL[f]
Is the QRS duration prolonged (>120 ms)?[b]	—	NULL[f]	NULL[f]
Is the corrected QT interval prolonged (>480 ms)?[b]	—	NULL[f]	NULL[f]
Is the troponin elevated (high-sensitivity cardiac troponin *t*>14 ng/L)?[b]	[26]	1.98	0.534

[a]LHR: likelihood ratio.

[b]Input for the Canadian Syncope Score (CSS).

[c]Included to prompt additional considerations.

[d]NA: not available.

[e]Artificially weighted for neurogenic loss of consciousness (LOC).

[f]Ratios not found in the literature.

[g]SBP: systolic blood pressure.

[h]Artificially weighted for orthostatic syncope.

Incorporation of an Evidence-Based Clinical Risk Tool and Recommendation Development

Based on the MISSION Syncope OptimalCare Pathway, recommendations included testing recommendations and disposition. Considering the lack of space on a mobile screen for large amounts of text, the Project MISSION team decided to categorize these recommendations into primary, secondary, and question specific. Primary recommendations were based on the user-selected DDx or that generated by the app if a user did not pick a diagnosis. The secondary recommendations would only be displayed if the app-generated DDx was discordant with the user-selected DDx. The question-specific recommendations referred to considerations specifically for orthostasis and neurogenic LOC if app users answered positively to only one of the questions associated with the etiology. Finally, based on the CSS classification, the implication of the risk score was provided as either discharge from the ED with outpatient workup (low risk), short-term observation of 6 hours or less (intermediate risk), or longer-term observation and admission (high risk).

User-Based Design and Technical Development

The app provides a presentation layer with limited logic and data storage and is designed to work online, with interactions processed on the backend web API and stored in the server

database. We collected all the information anonymously and recorded answers to all the assessment questions, the highest-ranked app-generated DDx, the user-selected DDx, and the CSS, allowing us to gather valuable information about syncope cases and to validate our model.

After input from frontline clinicians, we designed four screens that would be part of the main app workflow and reflect the steps in a real-world patient evaluation: Assessment, Differential Diagnosis, Risk Stratification, and Recommendations. See Multimedia Appendices 1 and 2 for the app installation guide and app assessment examples, respectively. The Assessment screen in the initial design consisted of a card deck–like UI, with each question on a separate card, a UI design most commonly seen in trivia apps. Our beta-testing focus group almost unanimously recommended against this design and instead recommended displaying the questions in list format. In addition to the aforementioned screens, based on user feedback, we determined that a home screen would be needed to have a place for the user to start new evaluations and view all the evaluations they have performed in the past. The list of evaluations included information about risks and the DDx of each. Based on the recommendations from the beta-testing focus group, we also included onboarding screens to the app. These are shown to the user when they start the app for the first time, and walk them through the purpose and usage of the app. In addition to the usability of the app, the beta- and alpha-testing phases allowed us to judge the consistency of the app diagnosis and recommendation.

Discussion

Principal Findings

Most CDS tools have a narrow scope of utility; they either are risk score calculators (eg, MDCalc) or provide information that might not be effectively specific (eg, ManageAnticoag) [20,27]. These latter apps often have a comprehensive list of variables, which limits the capability of providing specific, actionable recommendations and complicates decision making. The existing CDS tools for syncope are limited and only provide a risk score without subsequent recommendations or diagnostic support. Mobile apps also suffer from such a narrow scope, either focusing on patient education/engagement or translating risk score calculators to a mobile platform. Most publications on mobile apps fall into one of three categories: (1) proposal or initial development of a medical mobile app, (2) specific considerations in the development of mobile apps (eg, security or usability), and (3) position papers to improve standardization or evaluate the efficacy of apps. Our study falls into the first category. However, most of these apps are designed for patient engagement and education or risk calculators with embedded guideline recommendations. Our app development study is distinct in that provides prognostic information based on the validated CSS, an evidence-based approach to the DDx, and recommendations consistent with ACC/AHA guidelines [7,9]. There are also limited CDS tools for the diagnosis and management of syncope specifically. Several key learning points from this development process include (1) handling gaps in the literature and lack of a pre-existing study with such a model, (2) clearly and deliberately defining the scope of the app, (3) finding a balance between perfection and usability, and (4) dealing with technical challenges inherent to building a medical app. It is imperative to consciously identify and accept the potential of uncertainty in this type of project, as gaps exist in published research on patients presenting with syncope. We felt it most appropriate to mirror a clinical approach to these gaps by using adjunct variables, applying values to variables that are consistent with clinical experience, and deciding alternative approaches to addressing high-risk variables. This cannot be attempted in isolation and needs to be thoroughly vetted and tested within an interdisciplinary team with clinical, statistical, and technical expertise. As such, the differentiating factor of such a team is successfully combining the various domains of knowledge into a cohesive approach.

Defining the scope of an app is a necessary part of any software implementation and requires a dedicated focus during the ideation phase of the app design. Otherwise, there is a risk of building extraneous details into an app that will not be clinically useful and will be distracting to the user. The aim was to create an app that is clear, concise, and consistent, while providing the user with specific and actionable information and not complicating the UX with cluttered and generic information. With the MISSON Syncope app, we took a more holistic implementation approach with a focused set of recommendations. This app benefited from having a well-developed MISSION Syncope OptimalCare Pathway. A care pathway is the synthesis of guideline recommendations into an optimal pattern of practice considering complex clinical features that are part of the DDx and workup. Therefore, in similar software implementation, development of a care pathway should be a critical first step in the process. The pathway can guide what questions would be asked and what recommendations would be provided; without this, the app can become a series of disjointed questions addressing guideline recommendations that could be, at times, discordant with one another.

Successful app implementation requires a deliberate pursuit of a practical formula instead of striving for a perfect mathematical model. Creation of a nuanced, comprehensive formula results in a large number of variables, and although more extensive and possibly more accurate, it might not be practical for use as a mobile app. Additionally, since a perfect formula is untenable, given the current data and knowledge, potential incorrect assessments without acknowledging uncertainties and probabilities may provoke users to devalue the app and result in less acceptance. Therefore, a natural design point would be deciding between a perfect mathematical formula and a practical formula. Additionally, the formula was limited by a lack of study data and the LHRs available in the current literature, especially for orthostatic and neurogenic syncope. Importantly, the data collected by the app can provide useful information about filling gaps in the evaluation of syncope and guide future modifications.

Finally, in developing the app, we had to make a few deliberate decisions about the technical buildout. Deciding the supported screen sizes was especially challenging when building a cross-platform app that runs on both iOS and Android devices.

With limited resources and a small development team, we decided to use the BrowserStack App Live testing platform to test the app on numerous real devices. Another challenge was deciding whether a user should be required to log into the app. To promote easy adoption and to secure the backend API, we developed an anonymous authentication system using the Firebase (Google LLC) mobile platform. This also helped in making our data collection inherently anonymous. One challenge that does not have a clear solution is determining the veracity of each evaluation—in other words, determining which evaluations are genuine and which are hypothetical scenarios that users might be entering to test out the app. One solution can be to disregard the first few evaluations for each user; another could be to allow the user to flag such hypothetical scenarios. Future versions of the app could incorporate technical changes based on use.

App development is an iterative process with opportunities for continuous improvement. Having gone through the development process, our team has become more informed about how to apply app development to other disease states and develop similar clinical support tools.

Limitations

The limitations of this study included a lack of app evaluation by a larger number of users. The app will likely benefit from evaluation with input from a larger number of syncope cases by comparing the physicians' DDx and app-derived DDx for these cases. To test and improve our statistical model, we provide the user with a ranked order of differentials and determine the accuracy of the app by comparing the app-generated DDx to the user-selected DDx, thus creating a feedback loop for continuous improvement. Another limitation is that although we set up alpha- and beta-testing groups, we did not have extensive user feedback on how well the UI works and where it can be improved. Future work would include conducting a comprehensive retrospective chart review study, where we will evaluate a larger number of syncope cases that are presented in the ED against the DDx, recommendation, and risk stratification provided by the app. In addition, we plan to set up a process to continuously update the model parameters based on new studies and LHRs in continued literature reviews.

Conclusion

The MISSION Syncope app development integrated the current literature and clinical expertise to provide an evidence-based DDx, a prognosis using a validated scoring system, and recommendations based on clinical guidelines. This app demonstrates the importance of using research literature in the development of a CDS tool and applying clinical experience to fill the gaps in available research. It is essential for a successful app to be deliberate in pursuing a practical clinical model instead of striving for a perfect mathematical model, given available published evidence. This hybrid methodology can be applied to similar CDS tool development.

Acknowledgments

This project was funded by the National Heart, Lung, and Blood Institute (grant no. 1U01HL143508-01) to authors JL, MW, and SSS. All authors have participated in the work and have reviewed and agree with the content of the manuscript.

Conflicts of Interest

The authors have no financial interests to disclose. The manuscript is entirely original and has not been copyrighted, published, submitted, or accepted for publication elsewhere.

Multimedia Appendix 1
QR code and app download instructions.
[PDF File (Adobe PDF File), 54 KB - jmir_v23i11e25192_app1.pdf]

Multimedia Appendix 2
Sample screenshots of the mobile app.
[PDF File (Adobe PDF File), 744 KB - jmir_v23i11e25192_app2.pdf]

References

1. Sun BC, Emond J, Camargo JC. Characteristics and admission patterns of patients presenting with syncope to U.S. emergency departments, 1992-2000. Acad Emerg Med 2004 Oct;11(10):1029-1034 [FREE Full text] [doi: 10.1197/j.aem.2004.05.032] [Medline: 15466144]
2. Shen WK, Decker WW, Smars PA, Goyal DG, Walker AE, Hodge DO, et al. Syncope Evaluation in the Emergency Department Study (SEEDS): a multidisciplinary approach to syncope management. Circulation 2004 Dec 14;110(24):3636-3645. [doi: 10.1161/01.cir.0000149236.92822.07]
3. Smyth S, Gupta V, Williams M, Cowley A, Sirrine M, Surratt H, et al. Identifying guideline-practice gaps to optimize evaluation and management for patients with syncope. Can J Cardiol 2021 Mar;37(3):500-503. [doi: 10.1016/j.cjca.2020.05.022] [Medline: 32447058]

4. Probst MA, Kanzaria HK, Gbedemah M, Richardson LD, Sun BC. National trends in resource utilization associated with ED visits for syncope. Am J Emerg Med 2015 Aug;33(8):998-1001 [FREE Full text] [doi: 10.1016/j.ajem.2015.04.030] [Medline: 25943042]
5. Huff JS, Decker WW, Quinn JV, Perron AD, Napoli AM, Peeters S, American College of Emergency Physicians Clinical Policies Subcommittee (Writing Committee) on Syncope. Clinical policy: critical issues in the evaluation and management of adult patients presenting to the emergency department with syncope. J Emerg Nurs 2007 Dec;33(6):e1-e17. [doi: 10.1016/j.jen.2007.04.009] [Medline: 18035161]
6. Canzoniero JV, Afshar E, Hedian H, Koch C, Morgan DJ. Unnecessary hospitalization and related harm for patients with low-risk syncope. JAMA Intern Med 2015 Jun 01;175(6):1065-1067. [doi: 10.1001/jamainternmed.2015.0945] [Medline: 25894921]
7. Shen WK, Sheldon R, Benditt D, Cohen M, Forman D, Goldberger Z, et al. 2017 ACC/AHA/HRS guideline for the evaluation and management of patients with syncope: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol 2017 Aug 01;70(5):e39-e110 [FREE Full text] [doi: 10.1016/j.jacc.2017.03.003] [Medline: 28286221]
8. Li J, Gupta V, Smyth SS, Cowley A, Du G, Sirrine M, et al. Value-based syncope evaluation and management: perspectives of health care professionals on readiness, barriers and enablers. Am J Emerg Med 2020 Sep;38(9):1867-1874 [FREE Full text] [doi: 10.1016/j.ajem.2020.05.029] [Medline: 32739858]
9. Thiruganasambandamoorthy V, Kwong K, Wells GA, Sivilotti ML, Mukarram M, Rowe BH, et al. Development of the Canadian Syncope Risk Score to predict serious adverse events after emergency department assessment of syncope. CMAJ 2016 Sep 06;188(12):E289-E298 [FREE Full text] [doi: 10.1503/cmaj.151469] [Medline: 27378464]
10. Colivicchi F, Ammirati F, Melina D, Guido V, Imperoli G, Santini M. Development and prospective validation of a risk stratification system for patients with syncope in the emergency department: the OESIL risk score. ACC Curr J Rev 2003 Sep;12(5):70-71. [doi: 10.1016/j.accreview.2003.08.093]
11. Quinn JV, Stiell IG, McDermott DA, Sellers KL, Kohn MA, Wells GA. Derivation of the San Francisco Syncope Rule to predict patients with short-term serious outcomes. Ann Emerg Med 2004 Feb;43(2):224-232. [doi: 10.1016/s0196-0644(03)00823-0]
12. Reed MJ, Newby DE, Coull AJ, Prescott RJ, Jacques KG, Gray AJ. The ROSE (risk stratification of syncope in the emergency department) study. J Am Coll Cardiol 2010 Feb 23;55(8):713-721 [FREE Full text] [doi: 10.1016/j.jacc.2009.09.049] [Medline: 20170806]
13. Del Rosso A, Ungar A, Maggi R, Giada F, Petix NR, De Santo T, et al. Clinical predictors of cardiac syncope at initial evaluation in patients referred urgently to a general hospital: the EGSYS score. Heart 2008 Dec;94(12):1620-1626. [doi: 10.1136/hrt.2008.143123] [Medline: 18519550]
14. Probst MA, Lin MP, Sze JJ, Hess EP, Breslin M, Frosch DL, et al. Shared decision making for syncope in the emergency department: a randomized controlled feasibility trial. Acad Emerg Med 2020 Sep 02;27(9):853-865 [FREE Full text] [doi: 10.1111/acem.13955] [Medline: 32147870]
15. Pfister R, Hagemeister J, Esser S, Hellmich M, Erdmann E, Schneider CA. NT-pro-BNP for diagnostic and prognostic evaluation in patients hospitalized for syncope. Int J Cardiol 2012 Mar 08;155(2):268-272. [doi: 10.1016/j.ijcard.2010.10.013] [Medline: 21056483]
16. Sheldon R, Rose S, Connolly S, Ritchie D, Koshman M, Frenneaux M. Diagnostic criteria for vasovagal syncope based on a quantitative history. Eur Heart J 2006 Feb;27(3):344-350. [doi: 10.1093/eurheartj/ehi584] [Medline: 16223744]
17. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: clinical application in day-to-day practice. Indian J Ophthalmol 2009;57(3):217-221 [FREE Full text] [doi: 10.4103/0301-4738.49397] [Medline: 19384017]
18. Thiruganasambandamoorthy V, Taljaard M, Stiell IG, Sivilotti MLA, Murray H, Vaidyanathan A, et al. Emergency department management of syncope: need for standardization and improved risk stratification. Intern Emerg Med 2015 Aug 28;10(5):619-627. [doi: 10.1007/s11739-015-1237-1] [Medline: 25918108]
19. Albassam OT, Redelmeier RJ, Shadowitz S, Husain AM, Simel D, Etchells EE. Did this patient have cardiac syncope?: the rational clinical examination systematic review. JAMA 2019 Jun 25;321(24):2448-2457. [doi: 10.1001/jama.2019.8001] [Medline: 31237649]
20. American College of Cardiology. ManageAntiCoag App. URL: https://www.acc.org/tools-and-practice-support/mobile-resources/features/manageanticoag [accessed 2021-10-12]
21. American College of Cardiology. Guideline Clinical App. URL: https://www.acc.org/tools-and-practice-support/mobile-resources/features/guideline-clinical-app [accessed 2021-10-12]
22. QxMD. Calculate by QxMD. URL: https://qxmd.com/calculate [accessed 2021-10-12]
23. du Fay de Lavallaz J, Badertscher P, Nestelberger T, Zimmermann T, Miró Ò, Salgado E, et al. B-type natriuretic peptides and cardiac troponins for diagnosis and risk-stratification of syncope. Circulation 2019 Feb 25:2403-2418. [doi: 10.1161/CIRCULATIONAHA.118.038358] [Medline: 30798615]
24. Galizia G, Abete P, Mussi C, Noro G, Morrione A, Langellotto A, et al. Role of early symptoms in assessment of syncope in elderly people: results from the Italian group for the study of syncope in the elderly. J Am Geriatr Soc 2009 Jan;57(1):18-23. [doi: 10.1111/j.1532-5415.2008.02070.x] [Medline: 19054186]

25. Del Rosso A, Alboni P, Brignole M, Menozzi C, Raviele A. Relation of clinical presentation of syncope to the age of patients. Am J Cardiol 2005 Nov 15;96(10):1431-1435. [doi: 10.1016/j.amjcard.2005.07.047] [Medline: 16275193]
26. Christ M, Geier F, Popp S, Singler K, Smolarsky A, Bertsch T, et al. Diagnostic and prognostic value of high-sensitivity cardiac troponin T in patients with syncope. Am J Med 2015 Feb;128(2):161-170.e1. [doi: 10.1016/j.amjmed.2014.09.021] [Medline: 25447619]
27. MDCalc. MDCalc App. URL: https://www.mdcalc.com/ [accessed 2021-10-12]

Abbreviations

ACC: American College of Cardiology
ACEP: American College of Emergency Physicians
ACS: acute coronary syndrome
AHA: American Heart Association
API: application programming interface
CAD: coronary artery disease
CDS: clinical decision support
CHF: congestive heart failure
CPG: clinical practice guideline
CSS: Canadian Syncope Score
DDx: differential diagnosis
EBP: evidence-based practice
ECG: electrocardiogram
ED: emergency department
EM: emergency medicine
HRS: Heart Rhythm Society
LHR: likelihood ratio
LOC: loss of consciousness
LVOT: left ventricular outflow tract
MISSION: Multilevel Implementation Strategy for Syncope optImal care thrOugh eNgagement
PCP: primary care provider
PO: per os
SAEM: Society for Academic Emergency Medicine
SBP: systolic blood pressure
SDM: shared decision making
TLOC: transient loss of consciousness
UI: user interface
UX: user experience

Edited by R Kukafka; submitted 22.10.20; peer-reviewed by O El-Gayar, S Rostam Niakan Kalhori, A Teles; comments to author 22.12.20; revised version received 05.02.21; accepted 21.09.21; published 16.11.21.

Please cite as:
Amin S, Gupta V, Du G, McMullen C, Sirrine M, Williams MV, Smyth SS, Chadha R, Stearley S, Li J
Developing and Demonstrating the Viability and Availability of the Multilevel Implementation Strategy for Syncope Optimal Care Through Engagement (MISSION) Syncope App: Evidence-Based Clinical Decision Support Tool
J Med Internet Res 2021;23(11):e25192
URL: https://www.jmir.org/2021/11/e25192
doi:10.2196/25192
PMID:34783669

Review

The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review

Zainab Jan[1], MS; Noor AI-Ansari[2], PhD; Osama Mousa[2], MS; Alaa Abd-alrazaq[2], PhD; Arfan Ahmed[2,3], PhD; Tanvir Alam[2], PhD; Mowafa Househ[2], PhD

[1]College of Health and Life Sciences, Hamad Bin Khalifa University, Qatar Foundation, Education City, Doha, Qatar

[2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Education City, Doha, Qatar

[3]Department of Psychiatry, Weill Cornell Medicine, Education City, Doha, Qatar

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation, Education City
AI Luqta St AI-Rayyan
Doha, 5825
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Bipolar disorder (BD) is the 10th most common cause of frailty in young individuals and has triggered morbidity and mortality worldwide. Patients with BD have a life expectancy 9 to 17 years lower than that of normal people. BD is a predominant mental disorder, but it can be misdiagnosed as depressive disorder, which leads to difficulties in treating affected patients. Approximately 60% of patients with BD are treated for depression. However, machine learning provides advanced skills and techniques for better diagnosis of BD.

Objective: This review aims to explore the machine learning algorithms used for the detection and diagnosis of bipolar disorder and its subtypes.

Methods: The study protocol adopted the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. We explored 3 databases, namely Google Scholar, ScienceDirect, and PubMed. To enhance the search, we performed backward screening of all the references of the included studies. Based on the predefined selection criteria, 2 levels of screening were performed: title and abstract review, and full review of the articles that met the inclusion criteria. Data extraction was performed independently by all investigators. To synthesize the extracted data, a narrative synthesis approach was followed.

Results: We retrieved 573 potential articles were from the 3 databases. After preprocessing and screening, only 33 articles that met our inclusion criteria were identified. The most commonly used data belonged to the clinical category (19, 58%). We identified different machine learning models used in the selected studies, including classification models (18, 55%), regression models (5, 16%), model-based clustering methods (2, 6%), natural language processing (1, 3%), clustering algorithms (1, 3%), and deep learning–based models (3, 9%). Magnetic resonance imaging data were most commonly used for classifying bipolar patients compared to other groups (11, 34%), whereas microarray expression data sets and genomic data were the least commonly used. The maximum ratio of accuracy was 98%, whereas the minimum accuracy range was 64%.

Conclusions: This scoping review provides an overview of recent studies based on machine learning models used to diagnose patients with BD regardless of their demographics or if they were compared to patients with psychiatric diagnoses. Further research can be conducted to provide clinical decision support in the health industry.

(J Med Internet Res 2021;23(11):e29749) doi:10.2196/29749

KEYWORDS

machine learning; bipolar disorder; diagnosis; support vector machine; clinical data; mental health; scoping review

Introduction

Background

Bipolar disorder (BD) is a predominant mental disorder that involves dramatic shifts in mood and temper. It is the 10th most common cause of frailty in young adults and affects approximately 1% to 5% of the overall population [1]. It is mostly initiated during emotional states caused by disturbances in thinking, ranging from extreme mania and excitement to severe depression [2]. An epidemiological survey reported that its prevalence is rapidly increasing every year [3]. BD is associated with an evidently higher early mortality [4]. Bipolar patients have unfortunate life situations because these patients have a lifetime 9 to 17 years lower than that of normal people [5]. Additionally, several studies from various countries including Denmark and the United Kingdom state that this mortality difference has continuously been increasing since the last decades [6]. Although the maximum number of death cases in BD are due to cardiovascular diseases and diabetes, some death cases are due to unnatural events. Suicide is also relatively predominant in the patients with BD [6]. Suicide rates in patients with BD are 10%-20% higher than in the general population [4]. This context demonstrates significant background knowledge on bipolar disorder.

To effectively comprehend BD conditions and stipulate better treatment, primary exposure to mental disorders is a crucial phase. Different from finding other long-lasting situations that depend on laboratory trials and statistical analysis, BD is stereotypically detected based on patients' self-statements in precise surveys planned for uncovering specific types of feelings, moods, and public relations [4]. Owing to the growing accessibility of information relating to patients' mental health levels, artificial intelligence (AI) and machine learning (ML) skills are proving useful for deepening our comprehension of mental health situations, and they are promising methods to support psychiatrists in making better clinical decisions and analyses [7]. In recent years, AI techniques have shown superior performance in countless data-rich implementation frameworks, including BP [8,9].

In a previous review, Diego et al [10] discussed the applications of ML algorithms in diagnosing BD. They focused on 5 main application domains of ML in BD: diagnosis, prognosis, treatment, data-driven phenotypes plus research, and clinical direction. In contrast, the current review aims to evaluate existing literature on the applications of ML in BD diagnosis. Moreover, in the current review, we only focused on the role of ML in diagnosing BD and its types, which has not been previously comprehensively reviewed in any other study. We also discuss the strengths and challenges associated with the present work, future research guidelines for spanning the breach among the applications of ML procedures and patient diagnosis.

Research Problem

BD is misdiagnosed as depressive disorder that leads to difficulties and delay in the treatment of affected patients [1]. Approximately 60% of patients with BD are looking for treatment of major depressive disorders [11]. According to a National Chinese Mental Health Survey report, while the incidence of BD in China increased by 4.5% within a 12-month period, the recognition rate of BD as a depressive disorder increased to 39.9% [12]. Hence, there is an urgent demand to diagnose BD correctly. Moreover, ML increasingly provides various advanced methods to diagnose BD at the individual level to achieve better clinical results [10]. Many scientists have used support vector machine (SVM) algorithms to build BD classification models using neuroimaging information to differentiate BD from major depression [13]. In Taiwan, scientists have designed prediction algorithms using random forests that calculate the genetic risk scores of BD [14]. However, based on all the evidence, it is necessary to provide a scoping review that focuses on all applications of ML for BD diagnosis. The current review aims to explore how ML algorithms are used for better diagnosis of BD.

Methods

Review Approach

The current scoping review was conducted to provide an understanding regarding the role of ML in diagnosing BD. A scoping review is an approach that is systematically executed to enable researchers to examine emerging evidence from available studies on a specific topic [15]. It is also helpful for identifying knowledge gaps in a given field [15]. This scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines recommended in 2016 [16].

Search Strategy

Search Sources

We conducted a systematic search in 3 electronic databases: PubMed, Google Scholar, and ScienceDirect. We searched for articles published between January 2016 and December 2021. The search was conducted between March 16 to March 20, 2021. The references lists of the included articles were reviewed to check for possible articles that could be included.

Search Terms

The search strategies applied differed depending on the nature of the databases chosen for the search and are given in Multimedia Appendix 1. For example, PubMed allows the application of limiters such as "humans" and "English" language articles. In addition, further search terms for BD were added as we uploaded the references of Medical Subject Headings (MeSH) in PubMed. Google Scholar and ScienceDirect limit the number of search terms. Therefore, some search terms were not used when searching in these 2 databases. The intervention terms identified were ("Artificial Intelligence*" OR "Deep Learning" OR "Machine Learning" OR "Natural Language Processing" OR neural network* OR "unsupervised learning" OR "supervised learning"). The disorder terms identified were ("Bipolar disorder" OR "Bipolar 1 Disorder" OR "Bipolar 2 Disorder" OR "bipolar mood disorder" OR "bipolar affective disorder" OR "Cyclothymic Disorder" OR Cyclothym* OR manic*. Regarding search terms related to studies' outcome, which was bipolar disorder diagnosis, the search terms used were (diagnos* OR recog* OR prognosis OR detect* OR screening*).

The articles obtained from the search were uploaded to the Rayyan intelligent review application (Rayyan Systems Inc) in an EndNote (Clarivate) format [17]. This application allows researchers to collaborate and review articles at easily and at a faster pace [17]. Reviewers can create individual or collaborative reviews and make decisions regarding including or excluding the articles independently [17]. We considered 2 aspects when determining the key terms to be used for the current scoping review, which were population and interventions. The population we considered comprised Individuals with or without any health condition regardless of their gender, age, and ethnicity. The interventions considered include the ML models and algorithms used for diagnosing BD. The search terms were selected based on several scoping and systematic reviews we encountered during the preliminary search phase in the databases specified above.

Study Eligibility Criteria

Articles met the inclusion criteria if they achieved the main objective, namely providing an overview on the role of ML in diagnosing BD. The criteria identified for the inclusion and exclusion phases are given in Textbox 1.

Textbox 1. Criteria for study selection.

Inclusion criteria

- Empirical studies
- Peer-reviewed articles, theses, dissertations, and reports
- No restrictions related to machine learning algorithms and models
- No restrictions on country of study
- English language
- No restrictions related to population
- Bipolar disorder

Exclusion criteria

- Newspapers, magazines, reviews, proposals, and posters
- Any language other than English
- Machine learning algorithms that do not detect bipolar disorder
- Nonhuman subjects

Study Selection

In the first phase, 3 researchers (NA, OM, and ZJ) screened the titles and abstracts of the retrieved articles in an independent manner. In the second phase, the reviewers went through the full text of the articles included from the first phase. The retrieved articles were uploaded to the Rayyan intelligent review application in an EndNote format [17]. Disagreements were discussed amongst the 3 reviewers and decisions were made via consensus.

Data Extraction

For data extraction, a form was developed to include all the different data considered for the scoping review such as the ML model, accuracy, and type of data used. A description of the data extraction fields is included in Multimedia Appendices 2 and 3. Data extraction was performed independently by the 3 reviewers (NA, OM, and ZJ) using and Microsoft Excel (Microsoft Corporation). Any disagreements regarding the extracted data were resolved via consensus. A summary of all the data extracted from included studies is given in Multimedia Appendices 4.

Data Synthesis

This scoping review follows a narrative synthesis approach to synthesize the extracted data of the studies that made it to the final phase of inclusion and exclusion. From this analysis, we included studies that used ML models to assess participants with BD compared with other psychiatric disorders and healthy controls. The studies were classified based on the ML model used to diagnose BD, whether the model was an existing one or a novel one, BD type, data used, accuracy of diagnosis, other statistical measures, and whether the data used were private (gathered by the researchers) or public (open-access data). We also summarized the characteristics of the selected articles. Furthermore, we categorized the ML models into 10 categories and identified the characteristics of the selected studies that fitted under each category for the diagnosis of BD.

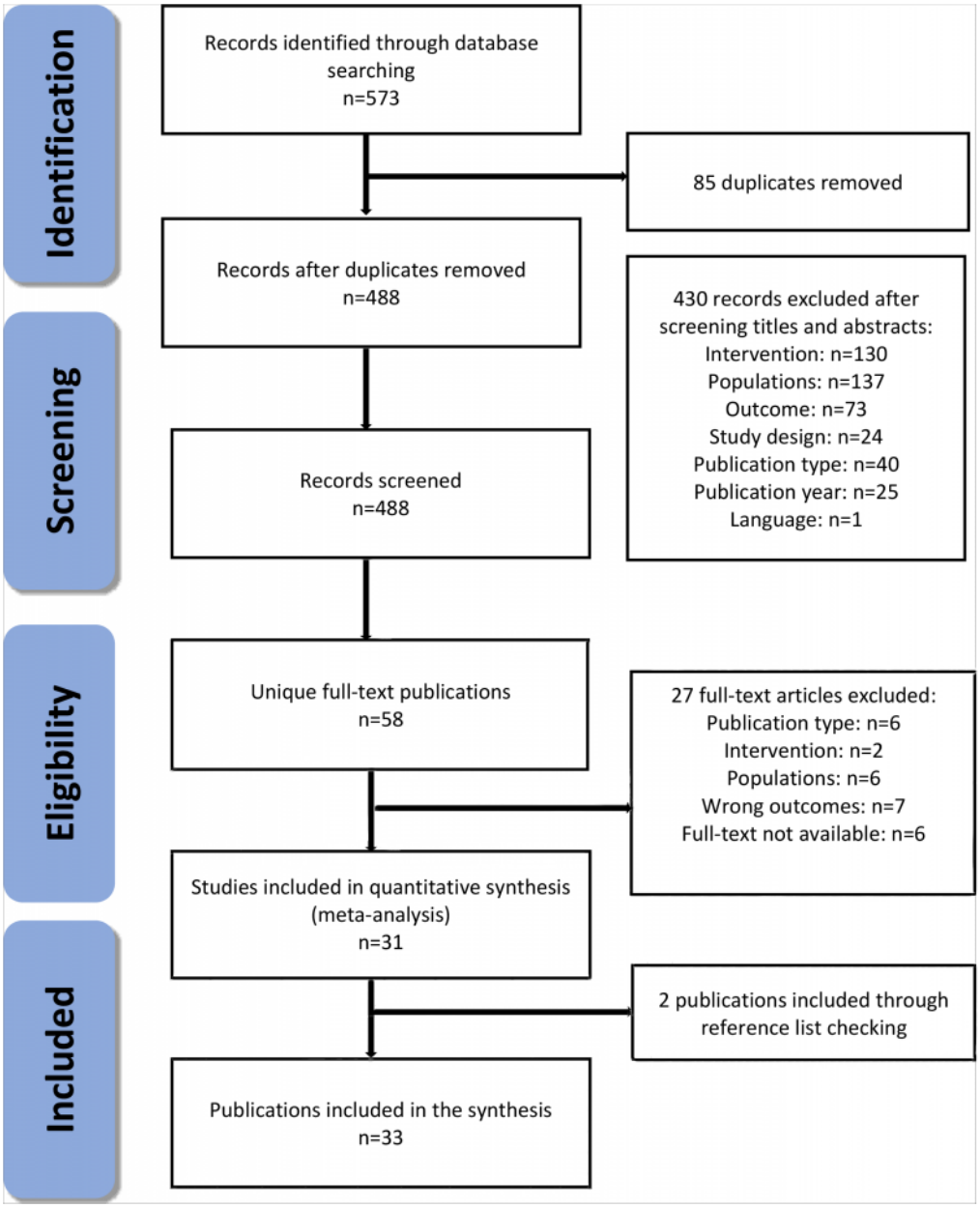
Results

Search Outcomes

In this scoping review, we retrieved 573 potential articles from 3 different databases and included 33 studies for data synthesis, as shown in Figure 1. Among these, 488 articles remained after eliminating 85 duplicates. In the first phase of screening the titles and abstracts of the articles, 430 records (wrong intervention=130 articles, population=137 articles, outcome=73 articles, study design=24 articles, publication types=40 articles, publication year=25 articles, and language=1 article) were excluded. In the second phase, we reviewed the full text of 58 articles and included 31 articles. Then, 2 additional studies were

added after checking the reference lists. Finally, 33 articles were selected for data synthesis.

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



Characteristics of the Included Studies

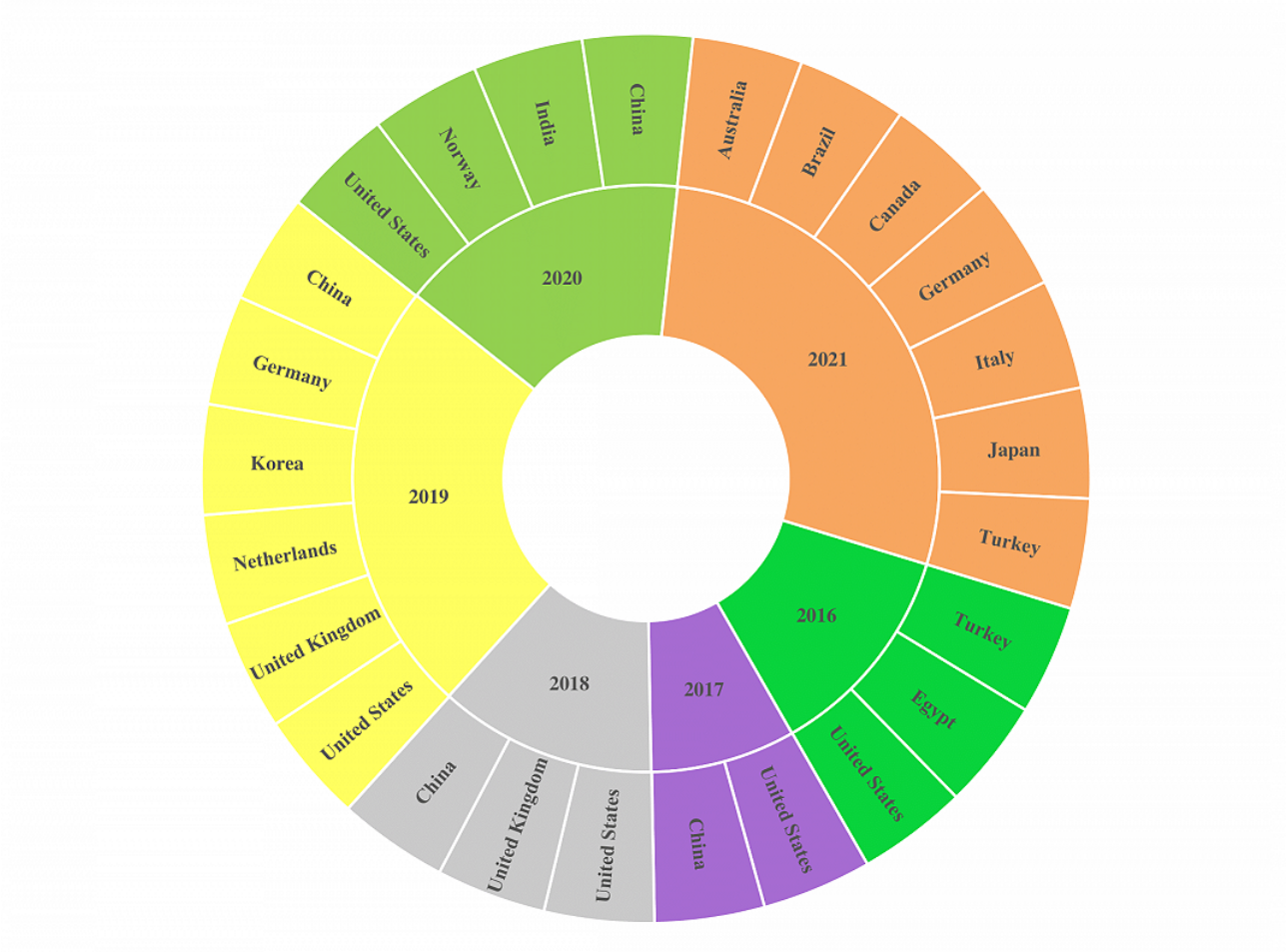
Among the 33 included articles, 30 were research articles (91%) [14,18-46], whereas 3 articles were conference proceedings (9%) [24,42,47], as shown in Table 1 and Multimedia Appendix 4. Articles were published in 14 different countries; China (8, 24%) [14,18-20,22,25,30,39], India (1, 3%) [21], Germany (2, 6%) [23,47], United Kingdom (1, 3%) [26], United States (8, 24%) [27,28,32,34,37,38,41,45], Korea (2, 6%) [29,36], Egypt (1, 3%) [31], Turkey (2, 6%) [31,43], Italy (1, 3%) [33], Brazil (1%) [47], Australia (1%) [35], the Netherlands (1, 3%) [36], Norway (1, 3%) [37], Canada (1, 3%) [40] and Japan (1, 3%) [46]; however, the highest numbers of articles were from China and the United States, as observed in Figure 2. The highest numbers of the articles were published in 2018 and 2019 (7, 21.21%), as shown in Figure 2 and Multimedia Appendices 5 and 6. The basic purposes of the included studies were model development (24, 73%), evaluation (5, 16%), data analysis (3, 9%), and model adaptation (2, 6%) for the diagnosis of BD. In the included studies, different types of BD were diagnosed using ML techniques such as BD type 1 (27, 81%), BD type 2 (27, 82%), psychotic bipolar (3, 9%), chronic bipolar (2, 6%), and first episode bipolar (1, 3%). Multimedia Appendix 4 provides the characteristics of the included studies, the purposes of the ML techniques used in these studies, and the types of ML models used to diagnose BD in the included studies.

Table 1. General characteristics of the included studies (N=33).

Characteristic	Studies, n (%)
Publication type	
Research articles	30 (91)
Conference proceedings	3 (9)
Publication status	
Published	33 (100)
Country of publication	
China	8 (24)
United States	7 (21)
United Kingdom	3 (9)
Canada	2 (6)
Germany	2 (6)
Brazil	1 (3)
Japan	1 (3)
Australia	1 (3)
Italy	1 (3)
Turkey	1 (3)
Korea	2 (6)
Norway	1 (3)
Netherlands	1 (3)
India	1 (3)
Egypt	1 (3)
Year of publication	
2021	6 (18)
2020	5 (15)
2019	7 (21)
2018	7 (21)
2017	3 (9)
2016	5 (15)
Basic purpose	
Model development	24 (73)
Evaluation	5 (15)
Data analysis	3 (9)
Model adaptation	2 (6)
Disorder type	
Bipolar disorder type 1	27 (82)
Bipolar disorder type 2	27 (82)
Psychotic bipolar	3 (9)
Chronic bipolar	2 (6)
First episode bipolar	1 (3)
Area	
Machine learning	33 (100)
Deep learning	3 (9)

Characteristic	Studies, n (%)
Purpose of machine learning algorithms	
Diagnosis and detection	33 (100)

Figure 2. Publications by year and country.



Types of ML Models Used to Diagnose Bipolar Disorder in the Included Studies

As shown in Multimedia Appendix 4, the included studies demonstrate 8 types of ML models that have been used to diagnose BD. The most common ML model used belonged to the classification model category, comprising 56% (18/33) of the studies [14,18-26,29-32,36-38,42]. The least commonly used models were natural language processing models [48], clustering algorithms [27], and deep learning–based models [29,30,32]; the various types of models and methods used in the included studies for diagnosing BD are presented in Table 2 and Multimedia Appendix 4.

Table 2. Machine learning models and algorithms, methods, and tools used in the included studies (N=33).[a,b]

Model categories	Number of studies, n (%)	Study ID
Classification models		
Support vector machine	9 (28)	[18-26]
Artificial neural network	4 (12.12)	[29-32]
Artificial neural network-particle swarm optimization	1 (3.03)	[31]
Random forest	4 (12.12)	[14,23,29,36]
Prediction rule ensembles	1 (3.03)	[35]
Gaussian process model	2 (6.06)	[37,38]
Nearest neighbor classification algorithm	1 (3.03)	[42]
Naive Bayes algorithm	1 (3.03)	[42]
Decision tree algorithm	1 (3.03)	[42]
Model-based clustering		
Growth mixture modeling	1 (3.03)	[41]
Linear discriminant analysis	1 (3.03)	[19]
Regression models		
Baseline logistic regression	1 (3.03)	[14]
Linear regression	3 (9.09)	[33,34,47]
Elastic net method	2 (6.06)	[33,47]
Least absolute shrinkage and selection operator	2 (6.06)	[19,34]
Fuzzy TOPSIS method	1 (3.03)	[39]
Clustering algorithms		
K-means clustering	1 (3.03)	[27]
Deep learning–based models		
Deep neural network	2 (6.06)	[29,30]
Convolutional neural network	1 (3.03)	[29]
DeepBipolar	1 (3.03)	[32]
Natural language–based model		
Natural language processing	1 (3.03)	[48]
Bipolar disorder assessment tools[c]		
Structured clinical interview for DSM-IV[d]	1 (3.03)	[33]

[a]Machine learning models/algorithms were not reported in 2 studies, of which 1 study used a novel machine learning approach to diagnose bipolar disorder type I. The name of the model is not mentioned.

[b]Machine learning methods were only reported in 8 studies.

[c]This is an interview-based assessment tool for diagnosis.

[d]DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition.

Classification Models

The includes studies employed 9 different types of classification models. In 9 (28%) of the 33 studies, SVM-based models were used to diagnose BD (specific types are not mentioned) [18-26]. In 1 study [18], this model was used to diagnose chronic BD and first-episode BD, whereas in 3 studies [19,21,26], SVM was used to diagnose type 1 and type 2 BD. However, SVM [24] was also used to diagnose unspecified types of BD. There are 4 studies (12%) that used artificial neural networks (ANNs) [29-32] for diagnosis purposes. Specifically, random forests were used in 4 studies (12%) [14,23,29,36] for diagnosing type 1 and type 2 BD, whereas in 2 studies (6%), Gaussian process models were used to diagnose BD type 1 [37,38]. ANN-particle swarm optimization (ANN-PSO) (3.03%) [31] was only used in 1 study to diagnose BD (types are not mentioned), whereas prediction rule ensembles (3.03%) [35], the decision tree algorithm (3.03%) [42], the nearest neighbor classification algorithm (K-NN) (3.03%) [42], and the naive Bayes algorithm (3.03%) [42] were employed to diagnose type 1 and type 2 BD.

Regression Models

The 33 included studies used 4 different types of regression models. Baseline logistic regression used in only 1 (3.03%) study for diagnosing BD and other psychiatric disorders [14]. Linear regression models were used in 3 (9.09%) studies [33,34,47] to diagnose type 1, type 2, and unspecified BD. In 2 (6.06%) studies [33,47], the elastic net method and least absolute shrinkage and selection operator (LASSO) [19,34] were used for diagnosing of type I, type II, and other unspecified BD types.

Model-Based Clustering

Linear discriminant analysis (LDA) and growth mixture modeling (GMM) were employed in 2 (6.06%) studies [19,41] for diagnosing type 1 and type 2 BD.

Deep Learning–Based Models

Among the 33 studies, 1 (3.03%) used deep neural networks and convolutional Neural Network algorithms [29], and 1 (3.03%) study employed DeepBipolar [32] to diagnose BD (types are not mentioned).

Natural Language–Based Model

A natural language processing model was employed by 1 (3.03%) study [48] to diagnose type 1 and type 2 BD.

BD Assessment Tools

Only 1 (3.03%) study [33] used SCID (Structured Clinical Interview for DSM-IV), a BD assessment tool, for diagnosing type 1 and type 2 BD.

Fuzzy TOPSIS Method

The Fuzzy TOPSIS method was employed in 1 (3.03%) study [39] for diagnosing type 1 and type 2 BD (3.03%).

Clustering Algorithms

In 1 study (3.03%) [27], K-means clustering was used for detecting psychotic BD.

Features of the Data Used in the Included Studies

The sample sizes were not consistent, and different sample sizes were used in the included articles ranging from 15 to 25,000. In 18 (56%) of the 33 studies, the sample size was less than 300 (56%), whereas in 12 (36.4%) studies, the sample size was above 300, as indicated in Table 3 and Multimedia Appendix 4. The most important feature of the included study was the data type. Multidimensional data were used in the selected articles, out of which data in 61.13% (19) of the studies belong to the clinical category, whereas 38.7 % (12) of the studies involved nonclinical data such as that in genomic and genome-wide association studies (GWAS). Private data sources (nongovernment sources or any other clinical data that are not publicly available) were the most commonly used in the included studies, whereas the least commonly used data sources were public (government sources, public databases, online websites, and freely available databases). Most of the included studies used already existing ML models for data evaluation (10, 30.3%), whereas the second common purpose was model adaptation (6, 18.2%). Only few studies developed novel ML models (2, 0.6%), as shown in Multimedia Appendix 4. The most common BD types mentioned in the selected studies were type 1 and type 2, whereas the least common types were chronic bipolar, first episode bipolar, and psychotic bipolar disorders, as observed in Table 1 and Multimedia Appendix 4.

Table 3. Features of data used in the included studies (N=33).

Feature	Value
Data set size (sample size),[a] n (%)	
<100	9 (28)
100-200	9 (28)
200-600	7 (21)
700-1000	3 (9)
>2000	2 (6)
Data type,[b] n (%)	
Clinical data	19 (58)
Nonclinical data	12 (36)
Data sources,[c] n (%)	
Private	21 (64)
Public	9 (28)
Sample type[d] (%)	
Disorder samples	>90
Healthy control	10

[a]Data set size was only reported in 30 studies.

[b]Data types were only mentioned in 31 studies. Clinical data include blood samples, electronic medical records, neurological data, magnetic resonance imaging data, electroencephalography and microarray expression data, whereas nonclinical data include phenotype data, genotype data, genomic data, and genome wide association studies.

[c]Public data include government sources, public databases, websites, and freely available databases, whereas private data include nongovernment sources, personal information, or data of specific hospitals or research organizations. Private data include databases that are not available in the public domain.

[d]More than 90% of the samples used in the included studies were bipolar disorder samples (regardless of type), whereas 10% of the samples were healthy control samples.

Types of Data Sets Used in the Included Studies

Data types were only mentioned in the 31 of the 33 studies. As shown in Table 4 and Multimedia Appendix 4, clinical and nonclinical data are reported in the included articles. The following data sets were collected from various public and private sources: clinical data (immune-inflammatory signature, blood sample, neuropsychological, neurocognitive, electroencephalography and PGBI-10M manic symptom data) [18-28,30,31,33,38,41-43,46,49] and nonclinical data (Cambridge Neuropsychological Test Automated Battery cognitive scores, microarray expression data sets, large-scale GWAS, fractional anisotropy, axial diffusivity, radial diffusivity, electronic medical records, bipolarity indices, affective disorder evaluation scale, daily mood ratings survey and phenotypic data sets) [14,19,21,24,29,32,34,36,37,45,47,49,50].

Table 4. Data set types used in the included studies (N=33).

Data type[a]	Study reference
Clinical data (n=19)	
Immune-inflammatory signature	[33]
Blood samples (serum)	[46]
Neuropsychological data	[18]
Neurocognitive data	[43]
Affective Disorder Evaluation scale	[19]
Magnetic resonance imaging (structural and functional)	[20-23,26-28,30,38,40]
Electroencephalography	[24,31]
PGBI-10M[b] manic symptom data	[41]
Microarray expression data set	[42]
Nonclinical data (n=12)	
CANTAB[c] cognitive scores	[34]
Large-scale genome-wide association	[14,45]
Phenotypic data set	[36,45,50]
Fractional anisotropy	[37]
Radial diffusivity	[37]
Axial diffusivity	[37]
Electronic medical record	[47]
Passive digital phenotypes	[36]
Bipolarity index	[19]
Daily mood ratings survey	[49]
Diffusion tensor images	[21,25]
Affective Disorder Evaluation scale	[19]
Activity monitoring	[29]
Genomic data	[31]

[a]In several studies, more than one data type was used.

[b]PGBI-10M: Parent General Behavior Inventory-10-Item Mania Scale.

[c]CANTAB: Cambridge Neuropsychological Test Automated Battery.

Statistical Validations of ML Models and Algorithms, Methods, and Tools Used in the Included Studies

The accuracies of the ML models and algorithms were reported in 24 studies, as shown in Table 5 and Multimedia Appendix 4. The accuracy level ranged from ≤70% to >91%. The accuracy level was ≤70% in 3 studies [25,32,36], 71%-78% in 7 studies [18,23,34,37,43,47,49], 83%-90% in 9 studies [14,20,21,26,28,29,31,33,42], and >91% in 5 studies [19,22,24,35,40]. The highest accuracy was 98%, found in only 1 study, whereas the lowest accuracy level was 64%. The mean value of the accuracy in 26 studies was 82.06%, whereas the median value was 84%.

Sensitivity was reported in only 15 studies; it ranged from ≤60% to >90%. Sensitivity was ≤60% in 1 study [46], 65%-67% in 2 studies [37,38], 75%-78% in 3 studies [14,18,47], 80%-88% in 8 studies [20-22,26,29,31,41,42], and above 90% in 1 study [35]. The mean value of sensitivity was 78.26%, whereas the median value was 82%. Moreover, specificity was only mentioned in 13 studies. The value of specificity ranged from ≤70% to 92% in 1 study [46], 74%-77% in 3 studies [18,42,49], 80%-90% in 6 studies [14,20,29,37,41,47], and >90% in 4 studies [21,22,26,35]. The mean specificity value was 85.36% and the median was 85.4%.

The proportion of the area under the curve (AUC) value was only reported in 10 studies, ranging from ≤69% to >97%. In 3 studies, the AUC ratio was ≤70% [23,36,45]; in 2 studies, it was 74%-78% [31,43] and 84%-88% [14,47], and in 3 studies [19,28,33], it was >90%. The maximum AUC ratio was 97%, whereas the minimum AUC value was 65%. The mean AUC value was 80.95% in 10 studies, whereas the median value was 81%.

Table 5. Statistical validation.

Statistics	Study reference
Accuracy, %[a]	
≤70	[25,31,36]
71-78	[18,23,34,37,43,47,49]
83-90	[14,20,21,26,28,29,31,33,42]
>91	[19,22,24,35,40]
Sensitivity, %[b]	
≤60	[46]
65-67	[37,38]
75-78	[14,18,47]
80-88	[20-22,26,29,31,41,43]
>90	[35]
Specificity, %[c]	
≤70	[46]
74-77	[18,38,43]
81-89	[14,20,29,37,41,47]
>92	[21,22,26,35]
AUC, %[d]	
≤70	[23,36,45]
74-78	[32,43]
84- 88	[14,47]
>91	[19,28,33]

[a]Ratio of accuracy was not reported in 7 studies. In some studies, different values were mentioned, so the overall values do not sum up.

[b]Sensitivity was not mentioned in 18 studies.

[c]Specificity was not mentioned in 20 studies.

[d]AUC: area under the curve. It is basically used for statistical validation of any data. AUC values were not reported in 23 studies.

Discussion

Principal Findings

Previous studies stressed the importance of ML classifiers to aid in diagnosing BD accurately, as it is frequently misdiagnosed. Approximately 60% of BD cases are misdiagnosed as major depressive disorders, and a proper diagnosis may take up to 10 years [46]. AI and ML exhibit considerable potential in clinical decision support and analysis with the help of big data, especially in mental health [7].

In this review, we explored the uses of ML techniques in diagnosing BD. From the 573 studies retrieved, 33 studies were included in this review. To explore the use of ML in diagnosing BD, the information was classified into 3 main categories as follows:

Machine Learning Models Used for Diagnosing BD

This review identified ML models, methods, and tools used for diagnosing BD, some of which did not use ML methods as the primary tool for diagnosis but used them as a supportive tool.

SVMs were the most commonly used ML models in diagnosing BD in 9 (27%) of the 33 studies, followed by ANNs (5, 15%), followed ensemble models (3, 9%), linear regression (3, 9%), and the Gaussian process model (2, 6%). Further, natural language processing, linear discriminant analysis, and logistic regression were used once in each study (3, 9%). Additionally, 7 studies applied other ML models that were emerging models or used a program to perform the diagnoses. However, only 1 study used a BD assessment tool, SCID, for the diagnosis of BD and an ML model as a supportive tool. Further, 1 study did not specify which ML model was employed. Hence, the use of ML models to diagnose BD is influenced by the diagnosis of BD, which is why studies have been exploring different ML models to better diagnose such mental disorders.

Data Sets Used in the Included Studies

The included studies used 2 types of data in diagnosing BD (clinical and nonclinical data). Clinical data were the most widely used, in 19 (53%) of the 33 studies. Among these 19 studies, 10 used magnetic resonance imaging (MRI) to classify bipolar patients compared to other groups. Other less commonly used data are mentioned in Table 4.

Nonclinical data were used in 12 studies (36%); some examples of nonclinical data used are large-scale GWAS (2, 6%), phenotypic data sets (2, 6%), diffusion tensor images (DTIs) (2, 6%) and other less commonly used data (Table 4). It is not surprising that nonclinical data are less commonly used because they mainly depend on surveys and tests related to mental disorders, which may lead to some biased results.

Validation of ML Models

The retrieved studies used 4 main validation measures to validate the ML models; these measures are accuracy, sensitivity, specificity, and AUC.

The accuracy of the ML models and algorithms was reported in 24 studies. The accuracy ranged from ≤70% to >91%. The highest accuracy achieved was 98% in only 1 study, whereas the lowest accuracy was 64%. Most studies achieved an accuracy of 83%-90% (9, 37.5%). The mean value of the accuracy was 82.06%. Moreover, sensitivity was only reported in 15 studies; it ranged from ≤60 to >90%. The mean value of sensitivity was 78.26%, whereas most studies (8, 53.3%) achieved sensitivity values between 80% and 88%. Furthermore, specificity was only mentioned in 13 studies. The value of specificity ranged from ≤70 to 92%. The mean value of Specificity was 85.4%, and most studies (6, 46.15%) achieved specificity values of 80%-90%. Finally, the AUC value was only reported in 10 studies, ranging from ≤69% to >97%. The maximum AUC value was 97%, whereas the minimum value was 65%. The mean AUC value was 81%. An important factor is that we were unable to compare the ML models and better categorize them owing to the variety of validation methods used in the reviewed studies. However, accuracy tended to be the most used measure in validating the ability of ML models to diagnose BD.

Comparison With Prior Work

Diego et al [10] conducted a systematic review that explored the applications of ML in diagnosing BD. The authors included articles from PubMed, Embase, and Web of Science published in any language up to 2017. They extracted 757 articles and included 51 studies in their review. They focused on categorizing the studies based on the data used to diagnose, treat, and prevent BD. Our focus was providing insight on the ML techniques used to diagnose various types of BD, including bipolar 1, bipolar 2, chronic bipolar, and episode bipolar. However, the articles lack information on the type of BD used to train and test the ML models (20 out of 33 studies did not specify the BD type). Thus, the data were categorized based on the ML model used to classify bipolar patients. Furthermore, we highlighted the advantages of the different data types used for different ML models. MRI data that were specifically used for SVMs and Gaussian process models showed good accuracy. However, EEG data used for SVMs showed higher accuracy than MRI data (98%), whereas DTI data showed lower accuracy than MRI and EEG data in case of SVMs (68.3%). Hence, we can infer that the predictive power and accuracy of ML models depend on the type of input data, as summarized in Table 6.

Table 6. Model performance metrics.

Data type	Study ID	Proposed model	Sensitivity, %	Specificity, %	Accuracy, %	AUC[a]
GWAS[b]	[14]	Random forest	77.7	85.4	85.2	NR[c]
Neuropsychological data	[18]	SVM[d]	76	77	77.0	NR
ADE[e] and BPx[f]	[19]	SVM	NR	NR	96.0	92.1
MRI[g]	[20]	SVM	85	85	85	NR
MRI	[21]	SVM	82.3	92.7	87.6	NR
MRI	[22]	SVM	87.5	97.1	92.4	NR
MRI	[23]	SVM	NR	NR	76.0	74
MRI	[26]	SVM	84.6	92.3	83.5	NR
MRI	[38]	Gaussian process model	66.4	74.2	70.3	NR
EEG[h]	[24]	SVM	NR	NR	98.0	NR
	[31]	ANN[i]	83.87	NR	89.89	NR
DTI[j]	[25]	SVM	NR	NR	68.3	NR
Activity monitoring	[29]	RF,[k] CNN,[l] and ANN	82	84	84	NR
Genomic data	[31]	ANN-PSO[m]	83.87	NR	89.89	NR
Immune-inflammatory signature	[33]	Linear regression and elastic net methods	NR	NR	86	97
EMR[n]	[47]	Linear regression and elastic net methods	75	81	78	84
CANTAB[o] cognitive score	[34]	Linear regression and LASSO[p]	NR	NR	71.0	NR
Phenotypic data set (passive digital phenotype)	[36]	RF	NR	NR	65	67
Fractional anisotropy, radial diffusivity, and axial diffusivity	[37]	Gaussian Process model	66.67	84.21	75.0	NR
PGBI-10M[q] manic symptom data	[41]	Growth mixture modeling	83	89	NR	NR

[a]AUC: area under the curve.
[b]GWAS: genome-wide association.
[c]NR: not reported in the article.
[d]SVM: support vector machine.
[e]ADE: Affective Disorder Evaluation.
[f]BPx: bipolarity index.
[g]MRI: magnetic resonance imaging.
[h]EEG: electroencephalography.
[i]ANN: artificial neural network.
[j]DTI: diffusion tensor images.
[k]RF: random forest.
[l]CNN: convolutional neural network.
[m]ANN-PSO: ANN-particle swarm optimization.
[n]EMR: electronic medical record.
[o]CANTAB: Cambridge Neuropsychological Test Automated Battery.
[p]LASSO: least absolute shrinkage and selection operator.
[q]PGBI-10M: Parent General Behavior Inventory-10-Item Mania Scale.

Future Research and Practical Implications

This review categorized the most common ML models and data used in diagnosing BD. Based on our findings, ML models can diagnose BD using clinical and nonclinical data. Future research should explore the studies involving patients in clinical and nonclinical settings to better evaluate the accuracy of the ML models.

Moreover, future studies should explore the influence of external factors like social media and the influence of the society on mental disorders to evaluate the influence of these factors on the patients and their effects on the performance of the ML models.

Furthermore, ML models should be compared with other traditional techniques for diagnosing BD like the Affective Disorder Evaluation (ADE) scale and Structured Clinical Interview for DSM-IV.

Only 2 studies reviewed used data sets with sizes above 2000, which is not surprising considering that most studies had data size as a limitation. In future studies, the ML models should be trained and validated on a larger data set and have a larger healthy control sample, as it was less than 10% in the reviewed studies.

As AI use in the health sector is growing rapidly, physicians should pay careful attention to some major issues that stand in the way of dealing with sensitive data such as medical information because of data ownership and security issues.

BD symptoms overlap with other mood disorders, specifically MDD, and this leads to the misdiagnosis of BD [20]. Future research should explore the main indicator that shows the patient is diagnosed with BD; for example, studies showed that patients diagnosed with BD have abnormal gray matter density in the MRI images of the brain. Another major indicator is regional homogeneity (ReHo), which indicates the activity of the brain while at rest [20,23]. Although some studies explored the use of some ML techniques that use binary classification methods such as SVMs and logistic regression, it is still not clear how ML techniques can distinguish BD, healthy people, and other mood disorders without the need for 2 groups (binary classification).

In addition, clinicians and researchers should explore the use of ML technology in clinical settings and address the clinical implications and outcomes of ML in diagnosing BD. Future investigations should focus on understanding of people's physiological and psychological behavior regarding the use of these technologies and the level of acceptance shown by physicians and patients. Finally, clinicians should explore the effectiveness of diagnosing models in clinical settings and develop predictive models that can predict mental disorders like BD.

Strengths and Limitations

Strengths

The present review was conducted to address the lack of scoping reviews that gather and categorize ML models used in diagnosing BD. The importance of this review stems from the fact that the traditional ways of diagnosing BD may lead to late diagnosis (an average of 10 years delay until formal diagnosis). This review explored studies that examined the ability of ML models to diagnose BD using a variety of data.

The most recent reviews focused on the implications of ML in patients with BD focused either on a specific ML model (neural networks) [51] or on the application of ML using MRI data [52]. This review explored the application of ML models in diagnosing BD without any limitations in terms of the technique or the type of data used, which gives a deeper insight into the technologies used in this field.

The studies considered in this review were the latest one to reduce bias in terms of date selection. We also conducted a backward referencing check by which we found 2 studies. Finally, the study selection included 3 reviewers working independently and any disagreements in the process were discussed and a decision was made upon consensus; this ensured reduced selection bias.

Limitations

This review included only 3 databases (PubMed, Google Scholar, and ScienceDirect), and other databases were not included, such as Embase, IEEE, Scopus, and the ACM Digital Library. This may have led to the absence of some studies that might be relevant to our review; for example, we did not include XGBoosting or LGBM, which are the most common ensemble models used for diagnosis purposes. Some of these databases were not included because of inaccessibility and time constraints. Moreover, we only considered articles published in the last 5 years (2016-2021). We missed categorizing supervised and unsupervised ML models, such as logistic regression, which is a supervised learning method.

We retrieved studies published in English only, which potentially led to the absence of other relevant studies published in other languages, especially French. Our study included data belonging to the United States, United Kingdom, China, Germany, Japan, Turkey, Korea, Italy, India, Canada, Norway, Egypt, Australia, Brazil, and the Netherlands. We missed including data from other populations. This made our results less comprehensive.

Furthermore, this review focused mainly on ML models diagnosing BD, regardless of what the patients were compared to in the training and testing sets (other psychiatric diagnoses) and regardless of the demographics of the patients. This may lead to biased decisions compared to other psychiatric diagnoses without having a healthy control sample. Moreover, our search queries lacked terms related to specific ML algorithms or models. Hence, we did not retrieve articles that used these terms in the title or abstract instead of ML. This again reduced the diversity of our scoping review.

Conclusions

This scoping review grouped recent studies based on the ML model used to diagnose patients with BD regardless of their demographics or their assessments compared to patients with other psychiatric diagnoses. We have also provided information about the data used and summarized the data that were most

commonly used in diagnosing BD. The goal of this review was to provide insights into how these technologies can help in faster and better diagnosis of BD and to promote their use in making clinical decisions in the health industry.

Acknowledgments

We thank the faculty of the Division of Information Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, for providing the opportunity to conduct this review.

Authors' Contributions

The review was developed under the supervision and guidance of MH, AA, and AAA. Each reviewer independently carried out the study selection and data extraction phase. NAA reviewed OM's work in both phases, OM revised ZJ's work, and ZJ revised NAA's work. Any disagreements with the decisions made were discussed and a decision was made upon consensus. All reviewers collaborated equally on the manuscript writeup and data extraction. TA helped with the classification of machine learning models as well as the designing of performance metrics. ZJ prepared the final manuscript file, and AAA and MH reviewed the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of queries used in various databases.
[DOCX File , 13 KB - jmir_v23i11e29749_app1.docx]

Multimedia Appendix 2
Description of data extraction fields.
[DOCX File , 13 KB - jmir_v23i11e29749_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies and purposes of machine learning techniques used in the studies.
[DOCX File , 17 KB - jmir_v23i11e29749_app3.docx]

Multimedia Appendix 4
Summary of all the data extracted from the included studies.
[DOCX File , 21 KB - jmir_v23i11e29749_app4.docx]

Multimedia Appendix 5
Fractions of articles by publication type.
[PNG File , 30 KB - jmir_v23i11e29749_app5.png]

Multimedia Appendix 6
Fractions of numbers of articles published by year.
[PNG File , 73 KB - jmir_v23i11e29749_app6.png]

References

1. Passos IC, Ballester PL, Barros RC, Librenza-Garcia D, Mwangi B, Birmaher B, et al. Machine learning and big data analytics in bipolar disorder: a position paper from the International Society for Bipolar Disorders Big Data Task Force. Bipolar Disord 2019 Nov 18;21(7):582-594. [doi: 10.1111/bdi.12828] [Medline: 31465619]
2. Marrie R, Walld R, Bolton J, Sareen J, Walker J, Patten S, CIHR Team in Defining the Burden and Managing the Effects of Psychiatric Comorbidity in Chronic Immunoinflammatory Disease. Increased incidence of psychiatric disorders in immune-mediated inflammatory disease. J Psychosom Res 2017 Oct 01;101:17-23 [FREE Full text] [doi: 10.1016/j.jpsychores.2017.07.015] [Medline: 28867419]
3. Dome P, Rihmer Z, Gonda X. Suicide risk in bipolar disorder: a brief review. Medicina (Kaunas) 2019 Jul 24;55(8):403 [FREE Full text] [doi: 10.3390/medicina55080403] [Medline: 31344941]
4. Navarro-Mateu F, Alonso J, Lim CCW, Saha S, Aguilar-Gaxiola S, Al-Hamzawi A, WHO World Mental Health Survey Collaborators. The association between psychotic experiences and disability: results from the WHO World Mental Health Surveys. Acta Psychiatr Scand 2017 Jul 25;136(1):74-84 [FREE Full text] [doi: 10.1111/acps.12749] [Medline: 28542726]

5. Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018 Mar;20(2):97-170 [FREE Full text] [doi: 10.1111/bdi.12609] [Medline: 29536616]
6. Staudt Hansen P, Frahm Laursen M, Grøntved S, Puggard Vogt Straszek S, Licht RW, Nielsen RE. Increasing mortality gap for patients diagnosed with bipolar disorder-A nationwide study with 20 years of follow-up. Bipolar Disord 2019 May 26;21(3):270-275. [doi: 10.1111/bdi.12684] [Medline: 30051555]
7. Malhi G, Bell E. Prepubertal bipolar disorder: a diagnostic quandary? Int J Bipolar Disord 2020 Apr 20;8(1):20 [FREE Full text] [doi: 10.1186/s40345-020-00187-0] [Medline: 32307595]
8. Lin E, Lin C, Lane H. Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. Int J Mol Sci 2020 Feb 01;21(3):969 [FREE Full text] [doi: 10.3390/ijms21030969] [Medline: 32024055]
9. Fernandes BS, Karmakar C, Tamouza R, Tran T, Yearwood J, Hamdani N, et al. Precision psychiatry with immunological and cognitive biomarkers: a multi-domain prediction for the diagnosis of bipolar disorder or schizophrenia using machine learning. Transl Psychiatry 2020 May 24;10(1):162-161 [FREE Full text] [doi: 10.1038/s41398-020-0836-4] [Medline: 32448868]
10. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: a systematic review. Neurosci Biobehav Rev 2017 Sep;80:538-554. [doi: 10.1016/j.neubiorev.2017.07.004] [Medline: 28728937]
11. Kessing LV, Andersen PK, Vinberg M. Risk of recurrence after a single manic or mixed episode - a systematic review and meta-analysis. Bipolar Disord 2018 Feb 14;20(1):9-17. [doi: 10.1111/bdi.12593] [Medline: 29239075]
12. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224 [FREE Full text] [doi: 10.1016/S2215-0366(18)30511-X]
13. Jie N, Zhu M, Ma X, Osuch E, Wammes M, Theberge J, et al. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans Auton Mental Dev 2015 Dec 26;7(4):320-331 [FREE Full text] [doi: 10.1109/tamd.2015.2440298]
14. Chuang L, Kuo P. Building a genetic risk model for bipolar disorder from genome-wide association data with random forest algorithm. Sci Rep 2017 Jan 03;7:39943 [FREE Full text] [doi: 10.1038/srep39943] [Medline: 28045094]
15. Munn Z, Peters M, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143-147 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
16. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02:467-473 [FREE Full text] [doi: 10.7326/m18-0850]
17. Rayyan Systems. 2021 Feb 30. URL: https://www.rayyan.ai/ [accessed 2021-03-02]
18. Sawalha J, Cao L, Chen J, Selvitella A, Liu Y, Yang C, et al. Individualized identification of first-episode bipolar disorder using machine learning and cognitive tests. J Affect Disord 2021 Mar 01;282:662-668 [FREE Full text] [doi: 10.1016/j.jad.2020.12.046] [Medline: 33445089]
19. Ma Y, Ji J, Huang Y, Gao H, Li Z, Dong W, et al. Implementing machine learning in bipolar diagnosis in China. Transl Psychiatry 2019 Nov 18;9(1):305-301 [FREE Full text] [doi: 10.1038/s41398-019-0638-8] [Medline: 31740657]
20. Li H, Cui L, Cao L, Zhang Y, Liu Y, Deng W, et al. Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques. BMC Psychiatry 2020 Oct 06;20(1):488-482 [FREE Full text] [doi: 10.1186/s12888-020-02886-5] [Medline: 33023515]
21. Achalia R, Sinha A, Jacob A, Achalia G, Kaginalkar V, Venkatasubramanian G, et al. A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder. Asian J Psychiatr 2020 Apr 01;50:101984. [doi: 10.1016/j.ajp.2020.101984] [Medline: 32143176]
22. Osuch E, Gao S, Wammes M, Théberge J, Willimason P, Neufeld RJ, et al. Complexity in mood disorder diagnosis: fMRI connectivity networks predicted medication-class of response in complex patients. Acta Psychiatr Scand 2018 Nov 06;138(5):472-482 [FREE Full text] [doi: 10.1111/acps.12945] [Medline: 30084192]
23. Schwarz E, Doan N, Pergola G, Westlye L, Kaufmann T, Wolfers T, IMAGEMEND Consortium, Karolinska Schizophrenia Project (KaSP) Consortium. Reproducible grey matter patterns index a multivariate, global alteration of brain structure in schizophrenia and bipolar disorder. Transl Psychiatry 2019 Jan 17;9(1):12-13 [FREE Full text] [doi: 10.1038/s41398-018-0225-4] [Medline: 30664633]
24. El Gohary MI, Al Zohairy TA, Eissa AM, El Deghaidy S, Hussein HM. An intelligent system for diagnosis of schizophrenia and bipolar diseases using support vector machine with different kernels. Int J Appl Eng Sci 2016 Oct 10:257568 [FREE Full text]
25. Deng F, Wang Y, Huang H, Niu M, Zhong S, Zhao L, et al. Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. Prog Neuropsychopharmacol Biol Psychiatry 2018 Feb 02;81:340-349. [doi: 10.1016/j.pnpbp.2017.09.006] [Medline: 28912043]

26. Frangou S, Dima D, Jogia J. Towards person-centered neuroimaging markers for resilience and vulnerability in bipolar disorder. Neuroimage 2017 Jan 15;145(Pt B):230-237 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.08.066] [Medline: 27622393]
27. Mothi SS, Sudarshan M, Tandon N, Tamminga C, Pearlson G, Sweeney J, et al. Machine learning improved classification of psychoses using clinical and biological stratification: update from the bipolar-schizophrenia network for intermediate phenotypes (B-SNIP). Schizophr Res 2019 Dec;214:60-69 [FREE Full text] [doi: 10.1016/j.schres.2018.04.037] [Medline: 29807804]
28. Liu Y, Mellem MS, Gonzalez H, Kollada M, Mahableshwarkar AR, Madrid A, et al. Highly predictive transdiagnostic features shared across schizophrenia, bipolar disorder, and adhd identified using a machine learning based approach. bioRxiv Preprint posted online on December 18, 2018. [doi: 10.1101/453951]
29. Jakobsen P, Garcia-Ceja E, Riegler M, Stabell L, Nordgreen T, Torresen J, et al. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. PLoS One 2020;15(8):e0231995 [FREE Full text] [doi: 10.1371/journal.pone.0231995] [Medline: 32833958]
30. Gong J, Chen G, Jia Y, Zhong S, Zhao L, Luo X, et al. Disrupted functional connectivity within the default mode network and salience network in unmedicated bipolar II disorder. Prog Neuropsychopharmacol Biol Psychiatry 2019 Jan 10;88:11-18. [doi: 10.1016/j.pnpbp.2018.06.012] [Medline: 29958116]
31. Erguzel T, Sayar G, Tarhan N. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Comput Appl 2015 Jun 18;27(6):1607-1616 [FREE Full text] [doi: 10.1007/s00521-015-1959-z]
32. Sundaram L, Bhat RR, Viswanath V, Li X. DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. Hum Mutat 2017 Sep 23;38(9):1217-1224 [FREE Full text] [doi: 10.1002/humu.23272] [Medline: 28600868]
33. Poletti S, Vai B, Mazza MG, Zanardi R, Lorenzi C, Calesella F, et al. A peripheral inflammatory signature discriminates bipolar from unipolar depression: a machine learning approach. Prog Neuropsychopharmacol Biol Psychiatry 2021 Mar 08;105:110136. [doi: 10.1016/j.pnpbp.2020.110136] [Medline: 33045321]
34. Wu M, Passos IC, Bauer IE, Lavagnino L, Cao B, Zunta-Soares GB, et al. Individualized identification of euthymic bipolar disorder using the Cambridge Neuropsychological Test Automated Battery (CANTAB) and machine learning. J Affect Disord 2016 Mar 01;192:219-225 [FREE Full text] [doi: 10.1016/j.jad.2015.12.053] [Medline: 26748737]
35. Parker G, Spoelma MJ, Tavella G, Alda M, Hajek T, Dunner DL, et al. Differentiating mania/hypomania from happiness using a machine learning analytic approach. J Affect Disord 2021 Feb 15;281:505-509. [doi: 10.1016/j.jad.2020.12.058] [Medline: 33387816]
36. Cho C, Lee T, Kim M, In HP, Kim L, Lee H. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4):e11029 [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
37. Linke JO, Adleman NE, Sarlls J, Ross A, Perlstein S, Frank HR, et al. White matter microstructure in pediatric bipolar disorder and disruptive mood dysregulation disorder. J Am Acad Child Adolesc Psychiatry 2020 Oct;59(10):1135-1145. [doi: 10.1016/j.jaac.2019.05.035] [Medline: 31330239]
38. Mwangi B, Wu M, Cao B, Passos IC, Lavagnino L, Keser Z, et al. Individualized prediction and clinical staging of bipolar disorders using neuroanatomical biomarkers. Biol Psychiatry Cogn Neurosci Neuroimaging 2016 Mar 01;1(2):186-194 [FREE Full text] [doi: 10.1016/j.bpsc.2016.01.001] [Medline: 27047994]
39. Han Y, Lu Z, Du Z, Luo Q, Chen S. A YinYang bipolar fuzzy cognitive TOPSIS method to bipolar disorder diagnosis. Comput Methods Programs Biomed 2018 May;158:1-10. [doi: 10.1016/j.cmpb.2018.02.004] [Medline: 29544775]
40. Palaniyappan L, Deshpande G, Lanka P, Rangaprakash D, Iwabuchi S, Francis S, et al. Effective connectivity within a triple network brain system discriminates schizophrenia spectrum disorders from psychotic bipolar disorder at the single-subject level. Schizophr Res 2019 Dec;214:24-33. [doi: 10.1016/j.schres.2018.01.006] [Medline: 29398207]
41. Jo B, Findling RL, Hastie TJ, Youngstrom EA, Wang C, Arnold LE, et al. Construction of longitudinal prediction targets using semisupervised learning. Stat Methods Med Res 2018 Sep 08;27(9):2674-2693 [FREE Full text] [doi: 10.1177/0962280216684163] [Medline: 28067113]
42. Saylan C, Yilancioglu K. Classification of schizophrenia and bipolar disorder by using machine learning algorithms. J Neurobehav Sci 2016;3(3):92. [doi: 10.5455/jnbs.1471026038]
43. Sonkurt HO, Altınöz AE, Çimen E, Köşger F, Öztürk G. The role of cognitive functions in the diagnosis of bipolar disorder: a machine learning model. Int J Med Inform 2021 Jan;145:104311. [doi: 10.1016/j.ijmedinf.2020.104311] [Medline: 33202371]
44. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 2018 May 07;14(1):91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]
45. Shafquat A, Crystal R, Mezey J. Identifying novel associations in GWAS by hierarchical Bayesian latent variable detection of differentially misclassified phenotypes. BMC Bioinformatics 2020 May 07;21(1):178-125 [FREE Full text] [doi: 10.1186/s12859-020-3387-z] [Medline: 32381021]
46. Idemoto K, Ishima T, Niitsu T, Hata T, Yoshida S, Hattori K, et al. Platelet-derived growth factor BB: a potential diagnostic blood biomarker for differentiating bipolar disorder from major depressive disorder. J Psychiatr Res 2021 Feb;134:48-56 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.12.051] [Medline: 33360224]

47. Suen PJC, Goerigk S, Razza LB, Padberg F, Passos IC, Brunoni AR. Classification of unipolar and bipolar depression using machine learning techniques. Psychiatry Res 2021 Jan;295:113624. [doi: 10.1016/j.psychres.2020.113624] [Medline: 33307387]
48. Chandran D, Robbins DA, Chang C, Shetty H, Sanyal J, Downs J, et al. Use of natural language processing to identify obsessive compulsive symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. Sci Rep 2019 Oct 02;9(1):14146-14147 [FREE Full text] [doi: 10.1038/s41598-019-49165-2] [Medline: 31578348]
49. Perez Arribas I, Goodwin GM, Geddes JR, Lyons T, Saunders KEA. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Dec 13;8(1):274-277 [FREE Full text] [doi: 10.1038/s41398-018-0334-0] [Medline: 30546013]
50. Fernandes BS, Karmakar C, Tamouza R, Tran T, Yearwood J, Hamdani N, et al. Precision psychiatry with immunological and cognitive biomarkers: a multi-domain prediction for the diagnosis of bipolar disorder or schizophrenia using machine learning. Transl Psychiatry 2020 May 24;10(1):162-163 [FREE Full text] [doi: 10.1038/s41398-020-0836-4] [Medline: 32448868]
51. Yasin S, Hussain SA, Aslan S, Raza I, Muzammel M, Othmani A. EEG based major depressive disorder and bipolar disorder detection using neural networks: a review. Comput Methods Programs Biomed 2021 Apr;202:106007. [doi: 10.1016/j.cmpb.2021.106007] [Medline: 33657466]
52. Kim Y, Na K. Application of machine learning classification for structural brain MRI in mood disorders: critical review from a clinical perspective. Prog Neuropsychopharmacol Biol Psychiatry 2018 Jan 03;80(Pt B):71-80. [doi: 10.1016/j.pnpbp.2017.06.024] [Medline: 28648568]

Abbreviations

AD: axial diffusivity
ADE: Affective Disorder Evaluation
AI: artificial intelligence
ANN-PSO: Artificial neural network-particle swarm optimization
BD: bipolar disorder
CANTAB: Cambridge Neurophysiological Test Automated Battery
DTI: diffusion tensor images
EEG: electroencephalography
EHR: electronic health record
FA: fractional anisotropy
fMRI: functional magnetic resonance imaging
GPC: Gaussian process classifier
GWA: genome-wide association data
LR: logistic regression
ML: machine learning
MRI: magnetic resonance imaging
NLP: natural language processing
OCD: obsessive compulsive disorder
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RD: radial diffusivity
RF: random forest
rs-fMRI: resting state functional magnetic resonance imaging
SVM: support vector machine
YBOCS: yellow-brown obsessive-compulsive disorder

Edited by R Kukafka; submitted 19.04.21; peer-reviewed by X Cheng, J Yang; comments to author 18.05.21; revised version received 02.07.21; accepted 27.07.21; published 19.11.21.

Please cite as:
Jan Z, AI-Ansari N, Mousa O, Abd-alrazaq A, Ahmed A, Alam T, Househ M
The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review
J Med Internet Res 2021;23(11):e29749
URL: https://www.jmir.org/2021/11/e29749
doi:10.2196/29749
PMID:34806996

Original Paper

Recruiting Participants for Population Health Intervention Research: Effectiveness and Costs of Recruitment Methods for a Cohort Study

Rania Wasfi[1,2,3], PhD; Zoe Poirier Stephens[2], MUP; Meridith Sones[4], MPH; Karen Laberee[4], MSc; Caitlin Pugh[4], MSc; Daniel Fuller[5], PhD; Meghan Winters[4], PhD; Yan Kestens[2,3], PhD

[1]Centre for Surveillance and Applied Research, Public Health Agency of Canada, Ottawa, ON, Canada

[2]Centre de recherche du CHUM, Université de Montréal, Montréal, QC, Canada

[3]École de Santé Publique, Université de Montréal, Montreal, QC, Canada

[4]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[5]School of Human Kinetics and Recreation, Memorial University of Newfoundland, St John's, NL, Canada

Corresponding Author:
Rania Wasfi, PhD
Centre for Surveillance and Applied Research
Public Health Agency of Canada
785 Carling Avenue
Ottawa, ON, K1A 0K9
Canada
Phone: 1 3435537356
Email: rania.wasfi@phac-aspc.gc.ca

Abstract

Background: Public health research studies often rely on population-based participation and draw on various recruitment methods to establish samples. Increasingly, researchers are turning to web-based recruitment tools. However, few studies detail traditional and web-based recruitment efforts in terms of costs and potential biases.

Objective: This study aims to report on and evaluate the cost-effectiveness, time effectiveness, and sociodemographic representation of diverse recruitment methods used to enroll participants in 3 cities of the Interventions, Research, and Action in Cities Team (INTERACT) study, a cohort study conducted in Canadian cities.

Methods: Over 2017 and 2018 in Vancouver, Saskatoon, and Montreal, the INTERACT study used the following recruitment methods: mailed letters, social media (including sponsored Facebook advertisements), news media, partner communications, snowball recruitment, in-person recruitment, and posters. Participation in the study involved answering web-based questionnaires (at minimum), activating a smartphone app to share sensor data, and wearing a device for mobility and physical activity monitoring. We describe sociodemographic characteristics by the recruitment method and analyze performance indicators, including cost, completion rate, and time effectiveness. Effectiveness included calculating cost per completer (ie, a participant who completed at least one questionnaire), the completion rate of a health questionnaire, and the delay between completion of eligibility and health questionnaires. Cost included producing materials (ie, printing costs), transmitting recruitment messages (ie, mailing list rental, postage, and sponsored Facebook posts charges), and staff time. In Montreal, the largest INTERACT sample, we modeled the number of daily recruits through generalized linear models accounting for the distributed lagged effects of recruitment campaigns.

Results: Overall, 1791 participants were recruited from 3 cities and completed at least one questionnaire: 318 in Vancouver, 315 in Saskatoon, and 1158 in Montreal. In all cities, most participants chose to participate fully (questionnaires, apps, and devices). The costs associated with a completed participant varied across recruitment methods and by city. Facebook advertisements generated the most recruits (n=687), at a cost of CAD $15.04 (US $11.57; including staff time) per completer. Mailed letters were the costliest, at CAD $108.30 (US $83.3) per completer but served to reach older participants. All methods resulted in a gender imbalance, with women participating more, specifically with social media. Partner newsletters resulted in the participation of younger adults and were cost-efficient (CAD $5.16 [US $3.97] per completer). A generalized linear model for daily Montreal recruitment identified 2-day lag effects on most recruitment methods, except for the snowball campaign (4 days), letters (15 days), and reminder cards (5 days).

Conclusions: This study presents comprehensive data on the costs, effectiveness, and bias of population recruitment in a cohort study in 3 Canadian cities. More comprehensive documentation and reporting of recruitment efforts across studies are needed to improve our capacity to conduct inclusive intervention research.

(J Med Internet Res 2021;23(11):e21142) doi:10.2196/21142

KEYWORDS

recruitment methods; Facebook recruitment; cost-effectiveness; built environment; intervention research; natural experiment; mobile phone

Introduction

Background

Urban interventions have the power to shape how people move, feel, and interact in cities, with the potential to improve health outcomes for all [1]. To understand the impacts of urban change on populations over time, researchers are using existing panel data sets [2-4] or collecting primary data [5,6]. Although representative population-based cohorts are key to successful population health intervention research [7], recruitment remains challenging [8,9]. Web-based recruitment strategies are increasingly used [10] because of their potential for wide reach over a short period and relatively low cost. Web-based technologies and related tools, such as smartphone apps or wearables, open new opportunities for data collection with lower participation burden. However, challenges to recruitment remain, including concerns over data privacy [11], time commitment for longitudinal studies [12], and limited reach toward marginalized populations [8,9]. All these can lead to biased samples, study delays, and increased costs [13].

Currently, few large-scale population health cohort studies have provided detailed reports on recruitment methods and effectiveness [6,14]. In a recent systematic review of studies that used Facebook to recruit participants in health, medical, or psychosocial research [10], only 19 out of 110 studies published between 2012 and 2017 reported details on cost and number of recruited participants by method. On average, the cost per completed participation through Facebook was CAD $6.79 (US $5.23; excluding staff time), although this varied widely (range CAD $1.36-$110 [US $1.05-$84.6]). Most of these studies were cross-sectional, with the exception of 2 cohort studies that focused on specific populations [15,16]. In a recent longitudinal web-based study examining physical activity through sustainable transport approaches in European cities, collaborations with local organizations, Facebook, mailing lists, and direct street recruitment were the most effective approaches to recruit participants, and Facebook was the most time-efficient method [6].

Objective

The overarching aim of this paper is to report and evaluate the effect of different recruitment methods used to enroll participants in a cohort study in 3 Canadian cities, led by the Interventions, Research, and Action in Cities Team (INTERACT) [5].

Methods

Study Design and Procedures

INTERACT uses a longitudinal design that is currently applied to 4 Canadian cities: Montreal, Saskatoon, Vancouver, and Victoria [5]. Local teams aimed to recruit approximately 300-person samples, except for Montreal, where the initial goal was 3000 participants across the Montreal region, where we aimed to evaluate a larger set of built environment interventions.

In our analyses, we only concentrated on 3 of the 4 INTERACT cities: Montreal, Saskatoon, and Vancouver, where we asked participants to report on how they had heard about the study. Interested participants were invited to complete a web-based eligibility questionnaire after consenting to the study. Participants could identify how they had heard about the study, either through a letter in the mail, referral from a friend or family member, social media (eg, Facebook and Twitter), met with study team, website, or other. In Saskatoon, they could choose from a few additional specific options (eg, posters on buses). This information was used to run the analyses by the recruitment method.

Eligible participants could choose from different levels of participation. The participants were asked to complete two web-based questionnaires: a health questionnaire and the Visualization, Evaluation, and Recording of Itineraries and Activity Spaces (VERITAS), a map-based activity space and social network questionnaire [17,18]. In addition, participants could choose to download and activate a smartphone app collecting GPS and accelerometer data for 30 days and answer Ecologic Momentary Assessment of well-being for 7 days. They could also choose to wear a hip-worn multi-sensor device (SenseDoc; Mobysens Technologies) for 10 days.

Target Sampling and Eligibility Criteria

Generally, participants were recruited through convenience sampling, with additional recruitment efforts aimed at reaching priority populations. Priority populations are those who are vulnerable or marginalized and need to be prioritized in research on healthy cities to ensure that every person has a fair and just opportunity to be as healthy as possible. Priority populations represent communities defined based on their age, gender, race, income, or ability. These priority populations include women, Black and Indigenous people, people with disabilities, people with low incomes, and older adults. For example, some social media campaigns specifically targeted underrepresented or low-income neighborhoods. The choice of inclusion criteria and survey questions was shaped by conversations with our knowledge user partners. Therefore, our recruitment approaches

varied based on each site's target sample and context. The specific recruitment tactics deployed in each city are described in the *Interventions and Participants by City* section. Inclusion criteria across all sites were as follows: being aged ≥18 years, being able to read or write English (or French in Montreal) well enough to answer a web-based questionnaire, and not planning to move out of the region in the next 2 years.

Interventions and Participants by City

In Vancouver, INTERACT evaluates the impact of the Arbutus Greenway, a 9-km former railway being redeveloped into a continuous walking and cycling corridor. Recruitment was conducted from April 20 to September 20, 2018 (123 days). The initial inclusion criteria required participants to live in one of the 8 forward sortation areas (FSAs) within 2 km of Greenway and be aged ≥45 years. To boost recruitment and to be consistent with age limits used in other sites, recruitment was then extended from June 18 to 12 FSAs within 3 km of the Arbutus Greenway and to adults aged ≥18 years. Participants were entered into a lottery to win one of 5 CAD $50 (US $39.5) Visa gift cards and a CAD $600 (US $461.5) gift certificate for a stay at a resort hotel.

In Saskatoon, INTERACT is studying the impact of a Bus Rapid Transit (BRT) system. Inclusion criteria included riding the bus at least once in a typical month or living within 800 m of the proposed BRT lines as determined by their postal codes. Recruitment ran from September 19, 2018, to January 4, 2019 (108 days). The participants received a CAD $10 (US $7.69) gift certificate upon completion of the health questionnaire. To encourage participants to contribute more data, participants were entered into a prize draw and received an additional chance of winning for each additional level of participation (VERITAS Questionnaire, app, or SenseDoc). Prizes included transit passes, a Bluetooth speaker, and headphones.

In Montreal, INTERACT evaluated the impacts of built environment interventions related to the Montreal Sustainability Plan (Plan Montréal durable 2016-2020). Interventions of interest include traffic calming measures, new transportation infrastructure, place-making, and greening programs. Target areas for recruitment included the Island of Montreal, Longueuil, Brossard, Saint-Lambert, and Laval. Participants were recruited between June 6 and December 21, 2018 (199 days). Participants were entered into a prize draw, with 20 CAD $100 (US $76.9) Visa gift cards and 1 prize with a value of CAD $500 (US $384.6): a choice of an iPad, a bicycle, or a stay at a resort hotel. Similar to Saskatoon, participants' chances of winning increased with their level of participation.

Recruitment Methods

Recruitment methods deployed at all sites included social media, news media, partner communications, snowball recruitment, and other methods, including in-person recruitment activities. Specific efforts and opportunities were tailored to each city.

Mailed Letters

Mailed letters were sent to Vancouver and Montreal. Mailing lists were rented from Canada Post. For the initial recruitment in Vancouver, 8614 personalized letters with an accompanying bookmark were sent to all homes in the 8 FSAs within 2 km of the Greenway where an individual aged ≥45 years lived. In Montreal, a mailed letter campaign with 3 types of options was sent to 15,000 people: a personalized letter with a postcard followed by a reminder postcard 2 weeks later (n=5000; group A), a personalized letter with a color card without a reminder (n=5000; group B), or a nonpersonalized postcard only (n=5000; group C). Letters were sent out by a third-party mail provider from the Canada Post Marketing program. Mailings were stratified by postal code to enable group identification based on the participants' postal code.

Social Media

All 3 cities used the INTERACT Twitter account (@teaminteract) and Facebook page [19] for promotion. In Montreal, the Centre de recherche du CHUM Facebook account also posted INTERACT content. In an effort to recruit underrepresented groups, messaging was adapted to younger people, and Facebook advertising was boosted in low-income postal codes in Montreal and Saskatoon. Facebook group administrators of community groups and nonprofit organizations in Montreal were contacted to post an invitation to the study.

News Media

Across all sites, the study was advertised through unpaid media coverage, through press releases to local media outlets, and by contact with journalists. In Montreal, the study was featured on news outlets such as *La Presse* and *Le Devoir*, CBC Montreal, *Montreal Gazette*, and TVA Nouvelles. In Saskatoon, the study was featured on CTV local news and CBC Saskatoon. In Vancouver, local CBC radio shows covered the study.

Newspaper Advertisement

In Montreal, information about the study was published in the Société de transport de Montréal section of the *Journal Métro*, free of charge.

Partner Communications

The research staff leveraged partner mailing lists, newsletters, and web-based spaces to promote the study. Efforts were made to reach community organizations working closely with citizens. Local teams also took advantage of institutional networks to share information, such as using listservs and university portals to advertise the study.

Snowball Recruitment

In Vancouver and Montreal, *Refer a friend* campaigns were launched using MailChimp. The participants were sent an email to share with a friend. Participants received a CAD $10 (US $7.69) gift card for every 2 referred friends who had signed up. In Saskatoon, participants were encouraged to share information about the study in their network, although no incentive was provided.

Other

We participated in a variety of community events to promote this study. In Saskatoon, research staff distributed flyers at bus terminals. In Vancouver, research staff attended farmers' markets, street parties, and seniors' activities around Arbutus Greenway. In Montreal, the team participated in city and

community events, distributed flyers at the Centre hospitalier de l'Université de Montréal, and visited local food banks. At these events, we collected email addresses for follow-up with interested people. All 3 cities designed and distributed posters to advertise the study. In Vancouver, posters were placed in cafés, local shops, and community spaces. In Saskatoon, posters were placed on buses. In Montreal, posters were placed in universities, community centers, and municipal buildings.

Recruitment Effectiveness Metrics

Cost

To calculate the cost of each recruitment method, we added the cost of producing materials (ie, printing costs), transmitting recruitment messages (ie, mailing list rental, postage, and sponsored Facebook post charges), and staff time. Staff time was assessed as 0.5 hours per Facebook post, 4 hours per in-person event, 2 hours per media publication, 2 hours per partner post, 50 hours for the mailed letters, and 35 hours for the snowball campaign. Compensation and expenses for prizes were excluded from the cost, as they were not consistent across sites.

Sociodemographic Profiles

We provide descriptive statistics on recruited populations for each method for age (4 categories: 18-34 years, 35-54 years, 55-64 years, and 65-88 years); gender (man, woman, and other); household income (CAD $0-49,999 [US $0-38,460]; CAD $50,000-99,999 [US $38,461-$76,922]; and CAD $100,000 [US $76,923] or more); education (less than a university degree, university degree, and graduate degree); and ethnicity (White; Indigenous or Aboriginal; and visible minorities, including South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, and Japanese).

Effectiveness

Recruitment method–specific effectiveness was determined by calculating the cost per completer, completion rate of the health questionnaire, and completion delay. The completion rate was calculated as the number of people who completed the health questionnaire divided by the number of eligible participants. Completion delay is defined as the time between the completion of eligibility and the health questionnaires.

Statistical Analyses

City differences in completion rate and completion delay were tested using the Kruskal–Wallis rank-based nonparametric method. A pairwise Wilcoxon test was used for multiple pairwise comparisons. Cost and compliance analyses per recruitment method were calculated for each city.

Modeling of daily recruitment by method and intensity was conducted for Montreal, where recruitment activities were recorded daily, and the sample size was larger. We modeled the number of participants recruited each day from the start to the end of the recruitment period (n=199 days). A recruited person was defined as someone who had completed the eligibility questionnaire and was deemed eligible and accepted to participate. Recruited participants were chosen over those who had completed the health surveys (eg, *completers*, above) to identify how different recruitment methods were able to reach participants and obtain their willingness to participate.

We fitted a distributed lag model using generalized linear regression to estimate the number of participants recruited on any given day. Predictive variables for each day were the type and intensity of recruitment campaigns, which included the following: (1) mailed letters, (2) people reached through paid Facebook posts and advertisements, (3) unpaid Facebook posts, (4) mailed reminders, (5) partner communications, (6) snowball recruitment campaigns, (7) wide-reach news media coverage (articles published in *La Presse* and *Le Devoir*, the 2 major francophone newspapers in Montreal), (8) smaller-reach news media coverage, and (9) other means of recruitment, including person events, posters, advertisements on web-based venues such as university websites, and classified advertisements. To consider the potential lag effect in recruitment for each method, we built different finite distributed lag weights ranging from 1 to 15 days [20]. Semilog transformations of the distributed lagged variables were used for Facebook reach:

$$y_d = \alpha + \beta_1 \times [x1(d\text{-}s)] + \beta_2 \times [\log x2(d\text{-}s)] + \beta_3 \times [x3(d\text{-}s)] + \beta_4 [x4(d\text{-}s)] + \beta_5 [x5(d\text{-}s)] + \beta_6 [x6(d\text{-}s)] + \beta_7 [x7(d\text{-}s)] + \beta_8 [x8(d\text{-}s)] + \beta_9 [x9(d\text{-}s)] + \mu_t$$

where y_d=number of participants recruited on a given day, lag length (s)=1, 2, 3....q, *x*1=mailed letters, *x*2=paid Facebook reach/1000, *x*3=unpaid Facebook posts, *x*4=mailed reminders, *x*5=partner communications, *x*6=snowball recruitment campaigns, *x*7=wide-reach news media coverage, *x*8=smaller-reach news media coverage, and *x*9=other means.

We retained the most efficient lag length (number of days) for each campaign type based on statistical significance, model fit (Akaike Information Criteria and Bayesian Information Criteria), and R-squared. We also visually examined the distribution of the residuals by plotting the observed and predicted estimates. Full details on the construction of the lagged variables and results of all combinations of different lag lengths (summarized in a CSV file) are provided on INTERACT's GitHub account [21]. Multimedia Appendix 1, Table S1 provides an example of 2- and 4-day lagged intensity variables. RStudio (version 3.6.1) was used to conduct all statistical analyses.

Results

The recruitment flowchart (Figure 1) provides details of recruitment, dropouts, eligibility, and completion of the health questionnaire. Participation choices by city are presented in Table 1.

Figure 1. Flowchart of recruitment numbers in the 3 Interventions, Research, and Action in Cities Team (INTERACT) cities.

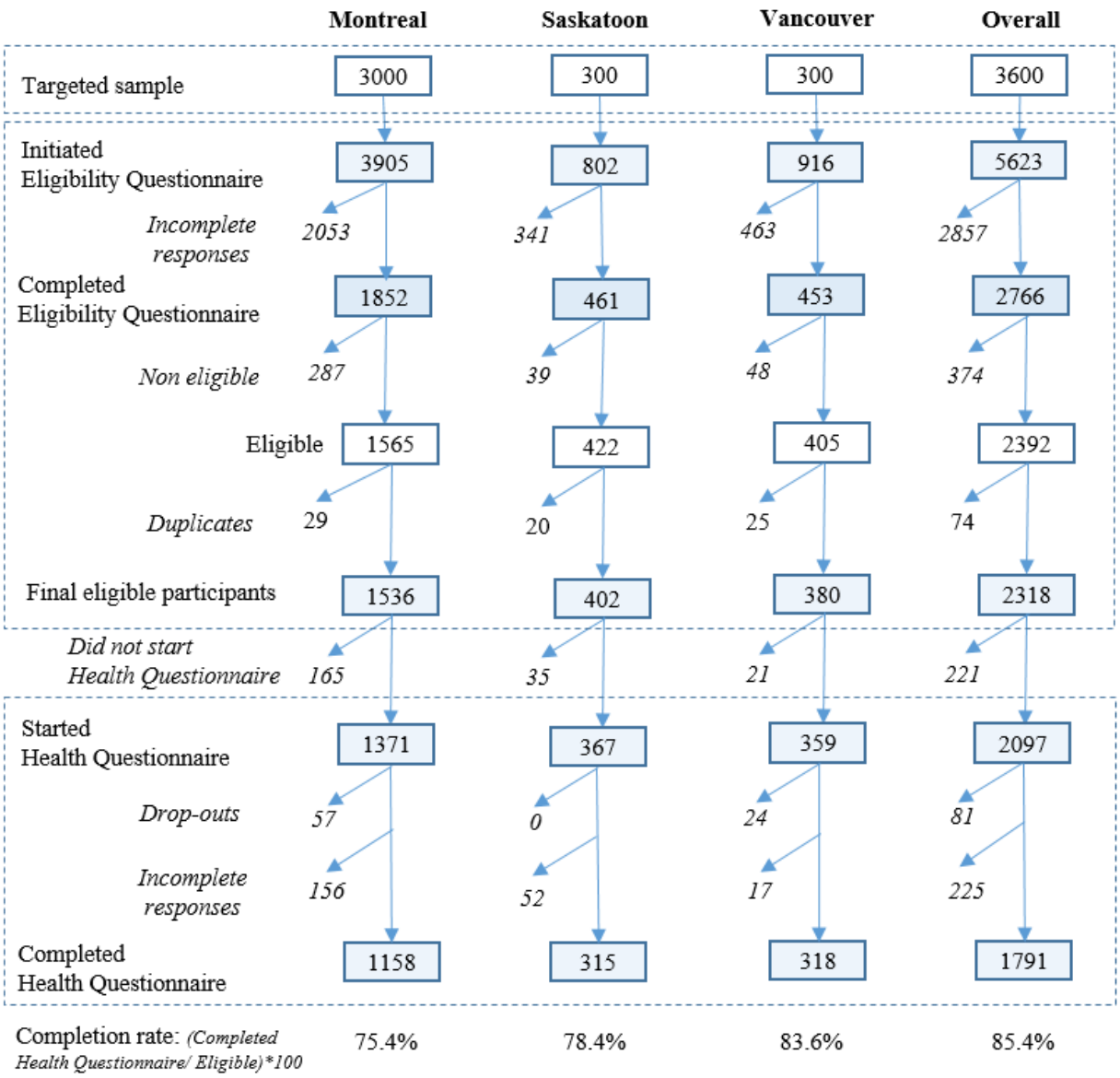


Table 1. Overall recruitment: participation option and health questionnaire completion by city[a].

	Montreal	Saskatoon	Vancouver
Total number of recruited participants per city, N (%)	1536 (100)	402 (100)	380 (100)
Total number of recruited participants who completed the health questionnaire per city, n/N (%)	1158/1536 (75.4)	315/402 (78.4)	318/380 (83.7)
Participation option			
1. Full participation (with a smartphone app and multi-sensor device)			
Number of recruited participants, N_1/N (%)	937/1536 (61)	225/402 (56)	161/380 (42.4)
Participants who completed a health questionnaire, n/N_1 (%)	744/937 (79.4)	179/225 (79.6)	134/161 (83.2)
2. Intermediate participation (with a smartphone app)			
Number of recruited participants, N_2/N (%)	277/1536 (18)	67/402 (16.7)	68/380 (17.9)
Participants who completed health questionnaire, n/N_2 (%)	201/277 (72.6)	56/67 (84)	51/68 (75)
3. Intermediate participation (with multi-sensor device)			
Number of recruited participants, N_3/N (%)	N/A[b]	13/402 (3.2)	72/380 (18.9)
Participants who completed health questionnaire, n/N_3 (%)	N/A	9/13 (69)	68/72 (94)
4. Basic participation (only questionnaires)			
Number of recruited participants, N_4/N (%)	322/1536 (20.9)	97/402 (24.1)	79/380 (20.8)
Participants who completed health questionnaire, n/N_4 (%)	213/322 (66.1)	71/97 (73.2)	65/79 (82.3)

[a]The percentage of participants who completed the health questionnaire is provided per participation option.

[b]N/A: not applicable.

Recruitment Methods and Corresponding Sociodemographic Profiles

Table 2 provides sociodemographic information using the recruitment method. City-specific numbers are provided in Multimedia Appendices 2 and 3. Most participants were recruited through social media (n=687). The participants were younger (mean age 41.8 years, SD 14.1, years) than those recruited through most other means, especially traditional media (mean age 58.5 years, SD 12.6, years). Other methods were more effective in recruiting younger participants, such as partner communications (mean 35.2 years, SD 16.1, years) or snowball sampling (mean 39.5 years, SD 15.1, years), compared with social media. Gender imbalance was strong across all methods, with 69% of all recruits identifying as women, 30% as men, and less than 1% as other genders. Social media recruitment was the most gendered (78% women vs 21% men vs 0.4% other) and letters the least (57% women vs 43% men). Recruits were distributed across income categories, with the highest share of lower-income participants (less than CAD $50,000 per year [US $38,461 per year]) recruited through partner communications (40%), *other* methods (32%), and social media (29%). All methods managed to recruit higher-income brackets (15.2% of the sample had household incomes equal to or above CAD $150,000 per year (US $115,384 per year), but this was particularly strong for mailed letters (22% of recruits by that category). Finally, most of the people in the sample identified as White (83.6%), followed by visible minorities (South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, and Japanese: 13.6%), and Indigenous (1.6%). There were differences in proportions among the cities: in Montreal and Vancouver, only 10.1% and 15.1% were visible minorities and 0.4% and 1.3% were Indigenous, whereas Saskatoon's sample consisted of 24.8% of visible minorities and 6.3% of Indigenous participants. Interestingly, in Saskatoon, 38.9% of the visible minority participants were recruited through partner communication.

Table 2. Demographic characteristics by recruitment method[a].

Demographics	Mailed letters (n=282)	Social media (n=687)	News media (n=230)	Partner communications (n=218)	Snowball recruitment (n=121)	Other (n=253)	Total (n=1791)
Age category (years), n (%)							
18-34	10 (3.5)	238 (34.6)	44 (19.1)	108 (49.5)	58 (47.9)	56 (22.1)	514 (28.7)
35-54	91 (32.3)	273 (39.7)	74 (32.2)	47 (21.6)	34 (28.1)	75 (29.6)	594 (33.2)
55-64	89 (31.6)	95 (13.8)	62 (27)	17 (7.8)	11 (9.1)	29 (11.5)	303 (16.9)
65-88	92 (32.6)	48 (7)	46 (20)	13 (6)	12 (9.9)	39 (15.4)	250 (13.9)
Education, n (%)							
Less than university degree	94 (33.3)	142 (20.7)	26 (11.3)	63 (28.9)	21 (17.4)	74 (29.2)	420 (23. 5)
University degree	85 (30.1)	259 (37.7)	90 (39.1)	79 (36.2)	51 (42.1)	82 (32.4)	646 (36.1)
Graduate degree	100 (35.5)	282 (41)	113 (49.1)	70 (32.1)	49 (40.5)	94 (37.2)	708 (39.5)
Gender, n (%)							
Male	120 (42.6)	144 (21)	84 (36.5)	67 (30.7)	45 (37.2)	81 (32)	541 (30.2)
Female	161 (57.1)	537 (78.2)	144 (62.6)	149 (68.3)	75 (62)	170 (67.2)	1236 (69)
Other	0 (0)	3 (0.4)	2 (0.9)	2 (0.9)	0 (0)	2 (0.8)	9 (0.5)
Income category, n (%)							
CAD $0- $49,999 (US $0- $38,460)	50 (17.7)	200 (29.1)	36 (15.7)	87 (39.9)	31 (25.6)	82 (32.4)	486 (27.1)
CAD $50,000- $99,999 (US $38,461-$76,922)	79 (28.0)	212 (30.9)	91 (39.6)	38 (17.4)	41 (33.9)	56 (22.1)	517 (28.9)
CAD $100,000-$149,999 (US $76,923-$115,383)	49 (17.4)	122 (17.8)	49 (21.3)	33 (15.1)	20 (16.5)	45 (17.8)	318 (17.8)
CAD $150,000-$199,999 (US $115,384-$153,845)	29 (10.3)	53 (7.7)	19 (8.3)	14 (6.4)	9 (7.4)	22 (8.7)	146 (8.2)
≥CAD $200,000 (US $153,846)	33 (11.7)	40 (5.8)	15 (6.5)	15 (6.9)	10 (8.3)	13 (5.1)	126 (7)
Ethnicity, n (%)							
White	246 (87.2)	591 (86.0)	217 (94.3)	155 (71.1)	93 (76.9)	195 (77.1)	1497 (83.6)
Indigenous or Aboriginal	<5 (0.7)	10 (1.5)	0	5 (2.3)	<5 (0.8)	11 (4.3)	29 (1.6)
Visible minorities	31 (11)	81 (11.8)	11 (4.8)	55 (25.2)	23 (19)	42 (16.6)	243 (13.6)

[a]Missing responses: age: 7.25% (130/1791); education: 0.95% (17/1791); gender: 0.28% (5/1791); income: 11.05% (198/1791); and ethnicity: 1.23% (22/1791).

Questionnaire Completion

Completion rate, calculated as the proportion of eligible recruits who completed the health questionnaire, varied by city and by recruitment method (Table 3). The completion rate was highest for Vancouver (83.6%) and lowest for Montreal (75.4%; Figure 1). The completion rate by recruitment method varied from 88.4% (mailed letters) to 72.5% (snowball recruitment), yet between-city variations were also observed. For example, Vancouver's completion rate for those recruited through letters was 97.1%, compared with 87.1% in Montreal. The time elapsed between eligibility and health questionnaire completion varied widely across participants and recruitment methods but did not differ between cities. Those recruited through letters were quickest to complete the questionnaires (mean 9.1 days, SD 29.9, days), whereas the slowest were those recruited through social media (mean 14.3 days, SD 37.7, days), followed by partner communications (mean 13.8 days, SD 31.4, days), media (mean 11.8 days, SD 34.6, days), and other means (mean 10.6 days, SD 25.8, days).

Table 3. Completion of eligibility and health questionnaires and time taken by recruitment method for baseline INTERACT[a] in Montreal, Saskatoon, and Vancouver.

	Recruitment method						
	Mailed letters	Social media	News media	Partner communications	Snowball recruitment	Other	Total
Number of participants who completed eligibility questionnaire (recruited), n (%)	319 (13.76)	944 (40.73)	284 (12.25)	264 (11.39)	167 (7.20)	340 (14.67)	2318 (100)
Number of participants who completed health questionnaire (completer), n (%)	282 (15.74)	687 (38.36)	230 (12.84)	218 (12.17)	121 (6.75)	253 (14.13)	1791 (100)
Average days from eligibility to completion of health questionnaire, mean (SD)	9.1 (29.9)	14.3 (37.7)	11.8 (34.6)	13.8 (31.4)	9.6 (28.6)	10.6 (25.8)	12.3 (10.4)
Completion rate, %	88.4	72.8	81.0	82.6	72.5	74.4	77.3

[a]INTERACT: Interventions, Research, and Action in Cities Team.

Cost-effectiveness

Cost per completer by recruitment method varied by city (Table 4). The average cost per completer for the 3 cities (Montreal, Saskatoon, and Vancouver) was CAD $23.28 (US $17.91). City-specific costs per completion were CAD $26.52 (US $20.4) in Montreal, CAD $23.80 (US $18.3) in Vancouver, and CAD $10.85 (US $8.35) in Saskatoon. Cost per completer by recruitment method varied by city. Partner communications was the most cost-effective recruitment method across cities, with an average cost of CAD $5.16 (US $3.97) per completer. They were particularly efficient in Saskatoon, costing <CAD $1(US $1.3) per completer. News media cost on average CAD $7.35 (US $5.65) per completer and generated a considerable number of participants in Montreal.

Social media, which generated the most recruits, came third in terms of cost-effectiveness across cities, at an average cost of CAD $15.04 (US $11.57) per completer. The highest recruitment cost resulted from mailed letters, at an average of CAD $108.30 (US $83.3) per completer (CAD $130.80 (US $100.6) in Montreal; CAD $83.56 (US $ 64.27) in Vancouver). Comparing different mailed options showed that personalized letters were much more cost-effective than postcards only, and reminder cards did not help recruitment. The cost per completer for group B (personalized letter and color card only; n=88) was CAD $60.11 (US $46.23), followed by group A (personalized letter, color card, and a reminder postcard; n=75) at CAD $106.68 (US $82.06), and group C (nonpersonalized postcard only) was the costliest at CAD $796.34 (US $612.57) per completer (n=8 recruitment).

Table 4. Cost per completer by city and recruitment method.

Reported recruitment method	Montreal (n=1158)	Saskatoon (n=315)	Vancouver (n=318)	Total (n=1791)
Mailed letters, n (%)[a]	148 (12.8)	0 (0)	134 (42.1)	282 (15.7)
Cost per completer, CAD$ (US$)[b]	130.80 (100.61)	N/A	83.56 (64.27)	108.30 (83.31)
Social media, n (%)	503 (43.4)	88 (27.9)	96 (30.2)	687 (38.4)
Cost per completer, CAD$ (US$)	13.35 (10.27)	16.13 (12.4)	22.91 (17.62)	15.04 (11.56)
News media, n (%)	226 (19.5)	4 (1.3)	0 (US 0)	230 (12.8)
Cost per completer, CAD$ (US$)	6.17 (4.75)	74.04 (56.95)	N/A	7.35 (5.65)
Other, n (%)	109 (9.4)	79 (25.1)	65 (20.4)	253 (14.1)
Cost per completer, CAD$ (US$)	18.05 (13.88)	21.60 (16.62)	72.70 (55.92)	33.20 (25.54)
Partner communications, n (%)	91 (7.9)	126 (40)	1 (0.3)	218 (12.2)
Cost per completer, CAD$ (US$)	7.32 (5.63)	0.88 (0.68)	347.10 (267)	5.16 (3.97)
Snowball recruitment, n (%)	81(7)	18 (5.7)	22 (6.9)	121 (6.8)
Cost per completer, CAD$ (US$)	16.40 (12.6)	0 (0)	59.80 (46)	21.85 (16.8)
Average cost per completer, CAD$ (US$)	26.52 (20.4)	10.85 (8.35)	23.80 (18.3)	23.28 (17.9)

[a]Number of participants who completed the health questionnaire Percentages indicate the proportion of city participants recruited through this specific method.

[b]Cost per completer includes the cost of all materials, expenses, and staff time and is expressed in Canadian dollars. Additional costs of participant compensation in Saskatoon (CAD $10 [US $7.69]) for questionnaire completion) are not included in this table.

Recruitment Modeling in Montreal

We modeled the number of people recruited per day over the 199-day recruitment period, which included 1536 participants from the Montreal cohort who completed the eligibility questionnaire and were willing to participate. The predictor variables included campaign events by recruitment type. Within the 199-day recruitment period, there were 227 campaign events, including (1) 151 days of paid Facebook posts and advertisements with an average reach of 2770 (SD 3558) potential participants (minimum=144, maximum=20,156); (2) 44 unpaid Facebook posts posted over 34 days; (3) a mailed letter campaign reaching 15,000 people; (4) a mailed reminder campaign reaching 5000 people; (5) 18 communications with partners who sent out newsletters or shared information on their web-based spaces; (6) 2 wide-reach media coverage events; (7) 6 smaller-reach media coverage events; (8) 1 snowball recruitment campaign; and (9) 16 other events, including 2 in-person community events.

The model performed relatively well overall, with an adjusted R-square of 0.78. The model parameters are listed in Table 5. Each coefficient should be interpreted as the effect of a campaign event or the number of participants recruited. The weights for each campaign are distributed over several days (specific lag per campaign type), and the sum of weights per campaign event equals 1. The model estimates 107 recruits that occur over 15 days for the letter campaign sent to 15,000 participants. Every 1% increase in Facebook reach per 1000 participants resulted in a 0.014 increase in recruitment. For example, an increase of 10% (277 participants) from the average Facebook reach of 2770 participants per day recruited an estimated 3.8 participants over 2 days. Unpaid Facebook posts recruited an estimated 3.2 participants over 2 days. On average, every wide-reach news article was associated with the recruitment of 164 participants over the course of 2 days. Finally, on average, each smaller-reach news media campaign was associated with an estimated recruitment of 10 participants. The best model fit and residual distribution indicated 2-day lag effects for all recruitment methods, except for 4 days for snowball recruitment, 15 days for the mailed letter campaign, and 5 days for the mailed reminder campaigns. Figure 2 shows the predicted and observed daily recruitment during the 199-day recruitment period.

Figure 2. Predicted and observed daily recruitment of Montreal's Interventions, Research, and Action in Cities Team (INTERACT) cohort.

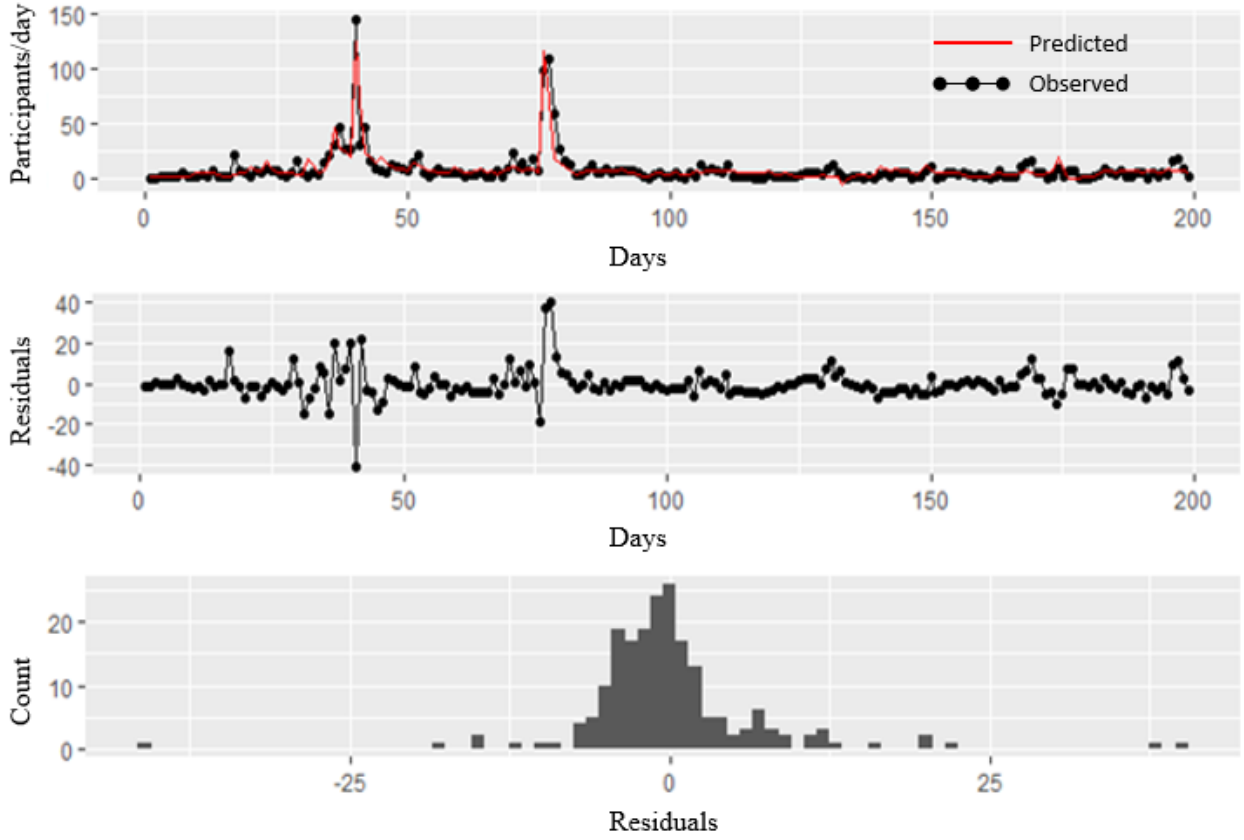


Table 5. Results of the regression model estimating the number of participants recruited per campaign event (number of observations=199)[a].

Predictors (lag in days)	Estimated number of recruited participants per campaign event	*P* value	95% CI
Mailed letters (15 days)	106.5[b]	.002	39.6 to 173.4
Mailed reminders (5 days)	16.9	.28	−13.5 to 47.2
Log paid Facebook reach per 1000 (2 days)	1.4[b]	<.001	0.6 to 2.1
Facebook unpaid posts (>2 days)	3.2[b]	.002	1.5 to 5.0
Wide-reach news media coverage (>2 days)	163.6[b]	<.001	149.0 to 178.2
Other recruitment means (>4 days)	−1.3	.05	−2.7 to 0.0
Partner communications (>2 days)	3.5	.16	−1.4 to 8.4
Snowball recruitment (>4 days)	2.5	.85	−24.4 to 29.3
Smaller-reach news media coverage (>2 days)	10.4[b]	.001	4.1 to 16.8
Intercept	4.3[b]	<.001	2.9 to 5.7

[a]R^2/R^2 adjusted=0.79/0.78.

[b]$P<.05$.

Discussion

Study Significance

This study documents the procedures and effectiveness of recruitment efforts to constitute baseline population-based samples in 3 Canadian cities as part of the INTERACT study. We further propose a daily recruitment modeling strategy that provides estimates of effectiveness for various recruitment campaigns, which we applied to the 199-day Montreal recruitment period. The existing literature generally lacks detailed reporting on recruitment performance [8,22]. To our knowledge, this is the first Canadian study to provide detailed performance indicators, including the time- and cost-effectiveness of different population-based recruitment methods.

Recruitment Method and Cost-effectiveness

Social media is a powerful and relatively cost-effective way to recruit participants. Approximately one-third of our participants were recruited through social media (687/1791, 38.4%), at an average cost of CAD $15.04 (US $11.56) per participant. However, these participants also took the longest to complete the questionnaires and had the lowest completion rates. This is possibly because social media users have direct access to the web-based recruitment material and therefore are more inclined to start the process, even with medium levels of motivation.

In contrast, letters had the highest and fastest completion rates (Table 6). Even if few of those receiving a letter were engaged, those who did were committed. Researchers recruiting samples for longitudinal studies or for studies requiring substantial time commitments from participants may want to consider such trade-offs and plan for potentially different follow-up rates by recruitment strategy. Previous studies found that follow-up rates were generally lower when participants were recruited on the web [23,24]. Although this study reports only baseline recruitment, future work should also consider differential attrition rates linked to the different recruiting methods.

The effectiveness of social media for recruitment has increased over the years. Montreal's (CAD $10.18 [US $7.83]) and Saskatoon's (CAD $14.45 [US $11.11]) social media costs (excluding staff time to facilitate comparisons with the literature) are in line with previously reported median costs of CAD $11.60 (US $8.92) for Facebook recruitment across 18 studies [10]. In Vancouver, where we initially targeted older adults living in a small geographic area, social media costs were higher (CAD $21.30 [US $16.38]) but in line with a recent Canadian study that recruited a hard-to-reach population through Facebook (CAD $19.27 [US $14.82]) [25].

Facebook posts were reported as an efficient recruitment method for a cohort study across 7 European cities (recruitment period: 2014-2016), although the cost per completer was not documented [6]. Earlier reports on one of the largest prospective cohort studies in the United Kingdom (recruitment period: 2009-2012) showed that Facebook posts are less efficient than mailed letters, SMS text messages sent on mobile phones, and emails [12]. However, since 2009, the share of the population with a social media account has grown, and Facebook has considerably refined its advertisement program, facilitating reach and recruitment [26]. Facebook advertisement features now make it possible to specifically target local areas or population segments based on individual profiles. These tools allow research teams to react to potential biases during the recruitment process, for example, by adjusting campaigns by targeting underrepresented geographic areas or population groups. Concomitantly, physical mail use has diminished, at least for letter correspondence. A previous study on smoking targeting young adults in Montreal (recruitment: 2011-2012) reported a 25% participation rate through letter recruitment [27], a much higher number than was achieved here (<1%). This difference might be linked to the presence of compensation and the age of the participants because CAD $10 (US $7.69) gift

certificates were given to those completing the survey. Transformations in communication habits and lower receptivity for mailed communication may also partly explain this difference. Finally, letters recruited older men than other methods, whereas social media recruited younger women, meaning these methods may be complementary.

Garnering attention for the study through newspaper articles was the second most effective strategy in Montreal, recruiting a high share of participants (226/1158, 19.5%) at a low cost (CAD $6.17 [US $4.75] per participant including staff time). Opportunities to publicize public health research in mainstream news outlets should be seized not only as a way to reach future participants, but also as a means to highlight existing research on the topic.

We recommend that researchers use multiple recruitment methods to amplify the impact of messaging and reach a greater diversity of participants. In Montreal, social media recruits were younger (mean age 41.8 years, SD 14.1, years) than those recruited through letters (mean age 58.8 years, SD 12.6, years) and media campaigns (mean age 51.4 years, SD 14.8, years). However, our social media recruitment profiles echo Canada's Facebook users: 38.2% (192/503) of our social media recruits were aged 18-34 years (42% of Facebook users in Canada), 40.2% (202/503) were aged 35-54 years (34% of Facebook users in Canada), 14.5% (73/503) were aged 55-64 years (12% of Facebook users in Canada), and 7.2% (36/503) were aged ≥65 years (10% of Facebook users in Canada) [28]. Gender imbalance was present across all recruitment methods but especially so among social media recruits: in Montreal, 77.1% (388/503) of the social media recruits were women compared with an average of 68% (788/1158) recruited through all other means of recruitment; in Saskatoon, 76.1% (67/88) were women compared with 73.7% (232/315); and in Vancouver, 85.4% (82/96) were women compared with 67.9% (216/318). Our gender imbalance is in the higher range of the 22 studies reporting a gender split in a 2016 systematic review, for which the median proportion of women was 61.1% [10]. We did not anticipate such a gender imbalance, although research has shown that women tend to join [29] web surveys and volunteer their time more than men [30], which may explain why more women completed the surveys.

INTERACT engaged with community organizations and institutions that had already established communication with citizens to promote the study. Low-income populations were best recruited through partner newsletters, consistent with previous research that supports working with community partners to reach priority populations [8,22]. Contacting citizens through such partners may improve the receptivity and trust of the participants [8]. This requires that the research team develop relationships with community partners who work directly with marginalized groups. Building relationships with both advocacy organizations as knowledge users and service delivery organizations as recruitment partners requires early and ongoing engagement from the research staff throughout the project.

When evaluating the extent of bias by sociodemographic factors in recruitment methods, one should consider the sociodemographic characteristics of each recruitment method. For example, because there is a higher share of female Facebook users, a nonbiased recruitment among Facebook users would result in more women participating. Similarly, mailing lists tend to have more up-to-date information on homeowners than tenants. This means that mail campaigns may be more effective in recruiting homeowners. Certain community organizations may have relationships with priority populations, facilitating recruitment. It is important to be aware of the sociodemographic population characteristics that these methods do reach before drawing conclusions about recruitment bias. Furthermore, although it is useful to assess bias for each specific method, using a variety of recruitment methods will tend to increase reach across sociodemographic groups. Table 6 presents a summary of the results and lessons learned from the INTERACT recruitment campaigns.

Table 6. Summary of strengths and weaknesses of each recruitment method, as seen in the INTERACT[a] study (lessons learned from INTERACT results).

Recruitment method	Strengths	Weaknesses
Mailed letters	• Highest and quickest completion rates • Effective at recruiting older populations • Higher share of older men than other methods	• Most expensive cost-per-completer rate
Social media	• Generally cost-effective for recruiting a large cohort • Effective in recruiting women and younger participants	• Had the lowest and slowest completion rates
News media	• Low cost-per-completer rate • Effective at recruiting older participants	• Low effectiveness for recruiting participants without a university degree • Little control from research team to garner attention from media
Partner communications	• High completion rate • Effective in reaching priority population participants • Effective for recruiting participants without a university degree • Least expensive cost-per-completer rate	• Slow completion rate • Important investments in time for building trust with partners
Snowball recruitment	• Ease of implementation through automated email campaigns	• Tends to reinforce trends within sample composition, because referred participants resemble their peers

[a]INTERACT: Interventions, Research, and Action in Cities Team.

Recruitment Method and Time Efficiency

One of the contributions of this study is that it provides a novel method to predict the number of daily recruits in a population-based recruitment effort, testing finite distributed lag weights for each recruitment approach. These results can inform the timing of different recruitment campaigns, including indications of their expected reach through time. We provide a detailed methodology, R syntax, and sample data on GitHub to facilitate the reproduction of this approach in other contexts. A systematic review [31] of modeling techniques used to predict recruitment to randomized clinical trials revealed a variety of modeling approaches, including Poisson and negative binomial models or Bayesian, time series, and Markov chain models. Using Poisson and negative binomial models does not capture the immediate rise in recruitment after special campaigns (eg, the peaks of recruited participants after wide-reach news media coverage). Bayesian, time series, and Markov chain analyses are less simple to reproduce [31]. With distributed lag weights as proposed in our study, ordinary least square models can be used [32]. Our model performed well in predicting daily recruitment, and recruitment-specific lags provided useful indications about temporal reach.

Limitations

The INTERACT study requires considerable time and effort from the participants. Beyond recruitment methods, the messaging used can affect diversity in recruitment. We used a variety of hooks and angles to capture the participants' attention. The impact of these factors was not assessed in this study. Moreover, differences in protocol in each city, notably compensation and prizes, may explain some of the variation in questionnaire completion rates among the cities. Future research may want to explore the impact of different types of messaging and visuals, including levels of participation and the impact of incentives on completion rates. It is possible that participants could have heard about the study from several sources, suggesting over- and underestimations and possible correction effects among methods. However, the model performs well in terms of cumulative recruitment; lag effects per method provide useful indications of the temporal dimensions of different recruitment approaches. Lower recruitment rates among priority populations are due to barriers such as distrust of participants and lack of knowledge in research, cultural beliefs and language issues [9], fear of stigmatization among those who may have engaged in high-risk behavior [8], issues related to low (technology) literacy, limited knowledge on the benefits the research might provide [33], privacy concerns, competing interests among busy participants [34], and lack of trust in web-based recruitment strategies [35]. The research team addressed these barriers in part by dedicating efforts to presenting the goals of the research and recruiting participants in person, connecting with community organizations that could promote the study among their clients and members, and providing phone or in-person assistance to participants answering questionnaires. Recruitment methods are only part of the equation for making participation more appealing and safer for all. Consequently, research teams should decide on the protocol at the outset and budget accordingly. Building trust and addressing logistical hurdles with priority populations are key goals for our next waves and should be considered at the forefront of any population health research.

Another limitation of the recruitment model is the inability to determine the sociodemographic profiles of the (unknown) exposed populations. Although Facebook Analytics provides profile statistics on the people reached through advertisements, such as sex, age, and geography, equivalent data for other recruitment methods were not available. For example, the number of people who are exposed to news media, snowball campaigns, or partner newsletters is unknown. Our model did not control for the demographic characteristics of those who were exposed to our campaigns.

Conclusions

Our study provides detailed documentation of recruitment efforts and the costs of population baseline samples across 3 Canadian cities. We also provide a novel lag-based modeling approach to evaluate the effectiveness of different recruitment strategies, illustrated using data from Montreal. Different recruitment methods had different costs, returns, and possible biases, suggesting that diversifying recruitment methods are useful to increase reach and sample diversity. Local contexts should not be ignored, as shown by the differences among the cities. Research teams should keep detailed logs of recruitment activities and ask participants to report how they were recruited to improve reporting of recruitment efficiency and costs. With increasing opportunities to collect large-scale citizen science data stemming from web-based platforms, smartphones, or wearables, setting up comprehensive recruitment strategies and better understanding how and why citizens choose to participate or not is important for the future of population-based research.

Acknowledgments

The authors would like to acknowledge all Interventions, Research, and Action in Cities Team (INTERACT) study national and local partners who helped with the recruitment efforts. The authors also thank Mr Benoit Thierry and Mr Alexandre Naud for revising parts of the R-code used in the analyses and Mrs Margot Gough for coordinating the recruitment efforts in Saskatoon. RW, the corresponding author, was affiliated with Université de Montréal as a postdoctoral fellow at the time this research was conducted and is currently affiliated with the Public Health Agency of Canada. The content and views expressed in this paper are those of the authors and do not necessarily reflect those of the Government of Canada.

INTERACT is funded under a Canadian Institutes of Health Research team grant: Environments and Health: Intersectoral Prevention Research/Subvention d'équipe: Santé et environnement: Recherche intersectorielle en prevention. July 13, 2016: Competition Code: 201607IP5. Funding reference number: IP2-1507071C.

Conflicts of Interest

YK, the nominated principal investigator holds shares in Mobysens Technologies Inc, a spin-off company that markets SenseDoc 2.0. SenseDoc is a multi-sensor device used for mobility (GPS) and physical activity (accelerometer) tracking in the Interventions, Research, and Action in Cities Team study. SenseDoc was filed as an invention in 2013 at Univalor, a valorization company affiliated with Université de Montréal and Centre de Recherche du CHUM. YK (nominated principal investigator) holds shares in Polygon Co, which markets the Visualization, Evaluation, and Recording of Itineraries and Activity Spaces tool referenced in this paper. the Visualization, Evaluation, and Recording of Itineraries and Activity Spaces was filed as an invention in 2012 at Univalor, a valorization company affiliated with Université de Montréal.

Multimedia Appendix 1
A hypothetical example to calculate the lagged intensity measure.
[DOCX File , 19 KB - jmir_v23i11e21142_app1.docx]

Multimedia Appendix 2
Number of recruited participants, rate, and time of completion by city and recruitment method.
[DOCX File , 21 KB - jmir_v23i11e21142_app2.docx]

Multimedia Appendix 3
Demographic characteristics by recruitment method and city.
[DOCX File , 26 KB - jmir_v23i11e21142_app3.docx]

References

1. Rydin Y, Bleahu A, Davies M, Dávila JD, Friel S, De Grandis G, et al. Shaping cities for health: complexity and the planning of urban environments in the 21st century. Lancet 2012 Jun;379(9831):2079-2108. [doi: 10.1016/s0140-6736(12)60435-8]
2. Wasfi R, Steinmetz-Wood M, Kestens Y. Place matters: a longitudinal analysis measuring the association between neighbourhood walkability and walking by age group and population center size in Canada. PLoS One 2017 Dec 19;12(12):e0189472 [FREE Full text] [doi: 10.1371/journal.pone.0189472] [Medline: 29261706]

3. Wasfi RA, Dasgupta K, Orpana H, Ross NA. Neighborhood walkability and body mass index trajectories: longitudinal study of Canadians. Am J Public Health 2016 May;106(5):934-940. [doi: 10.2105/ajph.2016.303096]
4. Fuller D, Pabayo R. The relationship between utilitarian walking, utilitarian cycling, and body mass index in a population based cohort study of adults: comparing random intercepts and fixed effects models. Prev Med 2014 Dec;69:261-266. [doi: 10.1016/j.ypmed.2014.10.022] [Medline: 25450496]
5. Kestens Y, Winters M, Fuller D, Bell S, Berscheid J, Brondeel R, et al. INTERACT: a comprehensive approach to assess urban form interventions through natural experiments. BMC Public Health 2019 Jan 10;19(1):51 [FREE Full text] [doi: 10.1186/s12889-018-6339-z] [Medline: 30630441]
6. Gaupp-Berghausen M, Raser E, Anaya-Boig E, Avila-Palencia I, de Nazelle A, Dons E, et al. Evaluation of different recruitment methods: longitudinal, web-based, pan-european physical activity through sustainable transport approaches (PASTA) project. J Med Internet Res 2019 Mar 09;21(5):e11492 [FREE Full text] [doi: 10.2196/11492] [Medline: 31066715]
7. Lavrakas PJ. Encyclopedia of Survey Research Methods. Thousand Oaks, CA: SAGE Publications; 2008.
8. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
9. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007 Jun 21;22(6):852-863 [FREE Full text] [doi: 10.1007/s11606-007-0126-3] [Medline: 17375358]
10. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
11. Roberts L, Sipes J. Ethical issues in online research. In: The Cambridge Handbook of Applied Psychological Ethics. New York, United States: Cambridge University Press; 2018:474-492.
12. Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to establish and follow up a large prospective Cohort study in the 21st century--lessons from UK COSMOS. PLoS One 2015 Jul 6;10(7):e0131521 [FREE Full text] [doi: 10.1371/journal.pone.0131521] [Medline: 26147611]
13. Allsworth JE. Invited commentary: recruiting for epidemiologic studies using social media. Am J Epidemiol 2015 May 15;181(10):747-749. [doi: 10.1093/aje/kwv007] [Medline: 25883156]
14. Nolte MT, Shauver MJ, Chung KC. Analysis of four recruitment methods for obtaining normative data through a Web-based questionnaire: a pilot study. Hand (N Y) 2015 Sep 14;10(3):529-534 [FREE Full text] [doi: 10.1007/s11552-014-9730-y] [Medline: 26330790]
15. King DB, O'Rourke N, DeLongis A. Social media recruitment and online data collection: a beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations. Can Psychol/Psychologie Canadienne 2014;55(4):240-249. [doi: 10.1037/a0038087]
16. Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Rio CD, et al. The comparability of men who have sex with men recruited from venue-time-space sampling and facebook: a Cohort study. JMIR Res Protoc 2014 Jul 17;3(3):e37 [FREE Full text] [doi: 10.2196/resprot.3342] [Medline: 25048694]
17. Kestens Y, Chaix B, Gerber P, Desprès M, Gauvin L, Klein O, et al. Understanding the role of contrasting urban contexts in healthy aging: an international cohort study using wearable sensor devices (the CURHA study protocol). BMC Geriatr 2016 May 05;16(1):96 [FREE Full text] [doi: 10.1186/s12877-016-0273-7] [Medline: 27151297]
18. Naud A, Sueur C, Chaix B, Kestens Y. Combining social network and activity space data for health research: tools and methods. Health Place 2020 Nov;66:102454 [FREE Full text] [doi: 10.1016/j.healthplace.2020.102454] [Medline: 33032243]
19. Team Interact. URL: https://www.facebook.com/interact.ca [accessed 2021-10-29]
20. Franses PH. Distributed lags. In: The New Palgrave Dictionary of Economics. London, U.K: Palgrave Macmillan; 2018:1-5.
21. TeamINTERACT / Recruitment_Wave1_Paper. GitHub. URL: https://github.com/TeamINTERACT/Recruitment_Wave1_Paper [accessed 2021-10-29]
22. Galli L, Knight R, Robertson S, Hoile E, Oladapo O, Francis D, et al. Using marketing theory to inform strategies for recruitment: a recruitment optimisation model and the txt2stop experience. Trials 2014 May 22;15(1):182 [FREE Full text] [doi: 10.1186/1745-6215-15-182] [Medline: 24886627]
23. Bajardi P, Paolotti D, Vespignani A, Eames K, Funk S, Edmunds WJ, et al. Association between recruitment methods and attrition in internet-based studies. PLoS One 2014 Dec 9;9(12):e114925 [FREE Full text] [doi: 10.1371/journal.pone.0114925] [Medline: 25490045]
24. Peels DA, Bolman C, Golsteijn RH, De Vries H, Mudde AN, van Stralen MM, et al. Differences in reach and attrition between web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012 Dec 17;14(6):e179 [FREE Full text] [doi: 10.2196/jmir.2229] [Medline: 23246790]
25. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. Internet Interv 2019 Sep;17:100246 [FREE Full text] [doi: 10.1016/j.invent.2019.100246] [Medline: 31080751]

26. Dodds L. Exclusive: Britons abandon Facebook as usage plummets by more than a third. The Telegraph. 2019. URL: https://www.telegraph.co.uk/technology/2019/07/06/exclusive-britons-abandon-facebook-usage-plummets-third/ [accessed 2021-10-16]
27. Frohlich K, Shareck M, Vallée J, Abel T, Agouri R, Cantinotti M, et al. Cohort profile: the Interdisciplinary Study of Inequalities in Smoking (ISIS). Int J Epidemiol 2017 Apr 01;46(2):e4. [doi: 10.1093/ije/dyv036] [Medline: 25948662]
28. Number of Facebook users in Canada from 2017 to 2026. Statista. 2021. URL: https://www.statista.com/statistics/282364/number-of-facebook-users-in-canada/ [accessed 2021-10-26]
29. Keusch F. Why do people participate in web surveys? Applying survey participation theory to internet survey data collection. Manag Rev Q 2015 Jan 9;65(3):183-216. [doi: 10.1007/s11301-014-0111-y]
30. Simmons WO, Emanuele R. Male‐female giving differentials: are women more altruistic? J Econ Stud 2007 Nov 06;34(6):534-550. [doi: 10.1108/01443580710830989]
31. Barnard KD, Dent L, Cook A. A systematic review of models to predict recruitment to multicentre clinical trials. BMC Med Res Methodol 2010 Jul 06;10(1):63 [FREE Full text] [doi: 10.1186/1471-2288-10-63] [Medline: 20604946]
32. Parker J. Linear distributed lag models. Reed College, Portland, Oregon. 2019. URL: https://www.reed.edu/economics/parker/312/online_slides/312_4-1.pdf [accessed 2021-10-16]
33. Mendoza-Vasconez AS, Linke S, Muñoz M, Pekmezi D, Ainsworth C, Cano M, et al. Promoting physical activity among underserved populations. Curr Sports Med Rep 2016;15(4):290-297. [doi: 10.1249/jsr.0000000000000276]
34. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Barriers and facilitators to patient and public engagement and recruitment to digital health interventions: protocol of a systematic review of qualitative studies. BMJ Open 2016 Sep 02;6(9):e010895 [FREE Full text] [doi: 10.1136/bmjopen-2015-010895] [Medline: 27591017]
35. Staffileno BA, Zschunke J, Weber M, Gross LE, Fogg L, Tangney CC. The feasibility of using Facebook, Craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2017;32(4):365-371. [doi: 10.1097/jcn.0000000000000360]

Abbreviations

BRT: Bus Rapid Transit
FSA: forward sortation area
INTERACT: Interventions, Research, and Action in Cities Team
VERITAS: Visualization, Evaluation, and Recording of Itineraries and Activity Spaces

Edited by R Kukafka, G Eysenbach; submitted 06.06.20; peer-reviewed by P Bajardi, D Gross; comments to author 08.08.20; revised version received 26.12.20; accepted 29.09.21; published 12.11.21.

Please cite as:
Wasfi R, Poirier Stephens Z, Sones M, Laberee K, Pugh C, Fuller D, Winters M, Kestens Y
Recruiting Participants for Population Health Intervention Research: Effectiveness and Costs of Recruitment Methods for a Cohort Study
J Med Internet Res 2021;23(11):e21142
URL: https://www.jmir.org/2021/11/e21142
doi:10.2196/21142
PMID:34587586

Original Paper

The Impact of Explanations on Layperson Trust in Artificial Intelligence–Driven Symptom Checker Apps: Experimental Study

Claire Woodcock[1], BA, MSc; Brent Mittelstadt[1], BA, MA, PhD; Dan Busbridge, MPhys, PhD; Grant Blank[1], BA, MA, PhD

[1]Oxford Internet Institute, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Claire Woodcock, BA, MSc
Oxford Internet Institute
University of Oxford
1 St Giles
Oxford, OX1 3JS
United Kingdom
Phone: 44 1865 287210
Email: cwoodcock.academic@gmail.com

Abstract

Background: Artificial intelligence (AI)–driven symptom checkers are available to millions of users globally and are advocated as a tool to deliver health care more efficiently. To achieve the promoted benefits of a symptom checker, laypeople must trust and subsequently follow its instructions. In AI, explanations are seen as a tool to communicate the rationale behind black-box decisions to encourage trust and adoption. However, the effectiveness of the types of explanations used in AI-driven symptom checkers has not yet been studied. Explanations can follow many forms, including *why*-explanations and *how*-explanations. Social theories suggest that *why*-explanations are better at communicating knowledge and cultivating trust among laypeople.

Objective: The aim of this study is to ascertain whether explanations provided by a symptom checker affect explanatory trust among laypeople and whether this trust is impacted by their existing knowledge of disease.

Methods: A cross-sectional survey of 750 healthy participants was conducted. The participants were shown a video of a chatbot simulation that resulted in the diagnosis of either a migraine or temporal arteritis, chosen for their differing levels of epidemiological prevalence. These diagnoses were accompanied by one of four types of explanations. Each explanation type was selected either because of its current use in symptom checkers or because it was informed by theories of contrastive explanation. Exploratory factor analysis of participants' responses followed by comparison-of-means tests were used to evaluate group differences in trust.

Results: Depending on the treatment group, two or three variables were generated, reflecting the prior knowledge and subsequent mental model that the participants held. When varying explanation type by disease, migraine was found to be nonsignificant (P=.65) and temporal arteritis, marginally significant (P=.09). Varying disease by explanation type resulted in statistical significance for input influence (P=.001), social proof (P=.049), and no explanation (P=.006), with counterfactual explanation (P=.053). The results suggest that trust in explanations is significantly affected by the disease being explained. When laypeople have existing knowledge of a disease, explanations have little impact on trust. Where the need for information is greater, different explanation types engender significantly different levels of trust. These results indicate that to be successful, symptom checkers need to tailor explanations to each user's specific question and discount the diseases that they may also be aware of.

Conclusions: System builders developing explanations for symptom-checking apps should consider the recipient's knowledge of a disease and tailor explanations to each user's specific need. Effort should be placed on generating explanations that are personalized to each user of a symptom checker to fully discount the diseases that they may be aware of and to close their information gap.

(J Med Internet Res 2021;23(11):e29386) doi:10.2196/29386

KEYWORDS

symptom checker; chatbot; artificial intelligence; explanations; trust; knowledge; clinical communication; mHealth; digital health; eHealth; conversational agent; virtual health care; symptoms; diagnostics; mobile phone

Introduction

Overview

Health care is a need so universal that the right to adequate medical care is enshrined in the Universal Declaration of Human Rights [1]. Yet, globally, governments face long-term challenges. High-income countries struggle with the financial burden of providing health care to aging populations with complex needs [2]. Meanwhile, half the world's population "lacks access to essential health services [3]..." More immediately, the COVID-19 pandemic is straining health care systems while making it necessary for care to be delivered remotely where possible [4].

These pressures have cultivated interest in "tools that use computer algorithms to help patients with self diagnosis or self triage," called symptom checkers (SCs) [5]. Although SCs have been developed using a broad range of techniques including Bayesian [6], rule-based, and deep learning methods [7], they are generally referred to as using artificial intelligence (AI) [8]. SCs are typically presented as smartphone chatbot apps and are created for one of two aims. First, akin to a visit to a primary care practitioner, some SCs are built to allow individuals to check what may be causing their health symptoms out of all common health conditions. Second, as a specific subset, some SCs check for symptoms of one disease only, typically COVID-19.

Private companies who build these apps [8] and government officials [9] believe SCs can improve provision of health care in two ways: (1) those with benign conditions can be easily triaged to less resource-intensive care, allowing human clinicians to focus on patients in need [10], and (2) SCs reduce the need for individuals to travel. This drives efficiencies in high-income countries [11] and allows those in countries with less care provision to access medical advice from remote locations.

Given their rapid global deployment, SCs have faced much scrutiny. Debate has focused on their accuracy, including their potential to fail to detect dangerous illnesses [12] or to provide overly cautious diagnoses [5,8]. Equally as important is the *human aspect* of SCs. By their nature, SCs abstract from the human interaction of a patient and physician [5]. The presented diagnosis may or may not provide additional insights as to why the app has come to a particular decision (Multimedia Appendix 1). Even assuming that an SC provides an accurate diagnosis, lay individuals must still follow triage instructions to achieve the intended benefits of SCs at a population level. Ensuring that users trust SCs is therefore of paramount importance if SCs are to achieve their intended benefits with regard to reducing the pressures faced by health care systems globally.

As with SCs, trustworthiness is seen as a necessary requirement for widespread adoption of AI. An essential component of trustworthiness is the capacity of a system (or its operator) to explain its behavior, for example, the rationale behind a particular diagnosis. Explanations are seen as a tool to communicate the rationale behind black box decisions to encourage user trust and adoption.

Recent studies have started to examine the effectiveness of different types of AI explanations [13-15], but to date no studies have specifically looked at SC explanations. Qualitative analysis of general sentiment toward medical conversational AI agents reveals a mixed reception [16], which suggests that choosing the right type of explanation in SCs is critically important for the systems to be well received. The lack of study of explanations in SCs poses challenges because poor explanations could reduce a person's inclination to use SCs, fuel health anxiety [17], or cause them to seek a second opinion from a human clinician, all of which would further burden health care systems. Furthermore, explanations can take many forms, including *why*-explanations and *how*-explanations, each of which may encourage user trust to different degrees. Social theories suggest that *why*-explanations are better at communicating knowledge and cultivating trust in laypeople, but this hypothesis has yet to be tested for SCs.

This paper presents the results of an exploratory study of layperson perception of SC explanations. Trust is used as a measurement of explanatory quality because good explanations are known to increase trust in AI systems [18], increasing the likelihood that the recipient will follow its output [19]. In this section, we begin by grounding the study in philosophical theories of contrastive explanation alongside cognitive psychology studies of causality. These theories emphasize that humans require explanations when information gaps are created. Although SCs currently address the need for explanation by explaining how the system derived the answer, humans typically prefer *why*-explanations. In the *Methods* section, we discuss the methodology used to conduct a study of 750 laypeople, where each participant was presented with a diagnosis of one of two diseases that was accompanied by one of four explanations. The results are presented in the *Results* section. In the *Discussion* section, we discuss findings that suggest trust may vary by explanation type for a lesser-known disease and that trust in explanation is significantly affected by the disease being explained. In the *Conclusions* section, we provide recommendations for SC system builders. The data and code required to reproduce all findings are publicly available [20].

Explanations in Theory and Practice

Overview

As this study focuses on the impact of varying types of SC explanations on layperson trust, it is first necessary to understand the purpose and use of explanations in AI. Explanations in theory and practice have long been studied by researchers, resulting in expansive literature on the subject. Here we adopt Lewis' [21] definition of an explanation as the provision of the causal history of an event. We also draw on Hilton [22] and Miller [23] to argue that explanations are ambiguous in being both a verb and a noun where an explainer *explains* (verb) *something* (noun) to *generate understanding* (verb) in a recipient.

Explanations and Their Purpose

Explanatory theory has typically focused on the function of explanations as a mechanism to transmit information about the *explanandum* (the event or phenomenon being explained) to

further inform a recipient's knowledge [24]. However, explanations can be used in many ways, including to persuade [25], assign blame [26], or even deceive the recipient [27,28]. Given these purposes, the explainer and the recipient may frequently have differing goals [25].

The goal of the SC explanation is to generate enough trust in the recipient so that they will follow triage instructions to reduce health system burden [8,10]. This is distinct from the goal of the recipient, which is to understand what is causing their symptoms [29].

Explanation Seeking and Knowledge

Explanations may be prevalent in human and social interactions, but they are not ubiquitous. People select when to seek an explanation [26]. Intrinsic provocation to seek an explanation, referred to as explanation seeking curiosity, is strongly predicted by future learning and future utility and moderately predicted by lack of knowledge [30].

The influential information gap theory described by Lowenstein [31] proposes that a delta between an individual's current knowledge and their desired knowledge cultivates curiosity. It is normatively correct to seek an explanation when an event or phenomenon does not fit one's mental model [32]. Empirically, information gaps have been shown to provoke explanation seeking in both children [33] and adults [34].

Knowledge and explanation are intimately linked [35]. When presented with information, the recipient must be aware of their lack of knowledge to seek an explanation [31]. An explanation must sufficiently transmit information [36], which is evaluated against prior beliefs and knowledge, particularly in knowledge-rich domains [25], the desired result being an updated mental model in the recipient [26]. The recipient's perception of an explanation subsequently provides insight into their knowledge of the phenomenon or the event being explained [22,37]. Hence, explanatory quality is critical because "explanations that are good and are satisfying to users enable users to develop a good mental model. In turn, their good mental model will enable them to develop appropriate trust in the AI [18]."

Everyday Explanations

Explanations can take many forms, for example, scientific, causal, teleological, or *everyday*. A universally accepted taxonomy does not exist [38]. Given this study's focus on laypeople and prior scholarship suggesting their utility in AI [23,38], our focus will be on everyday explanations. These are a form of explanation, which are commonly observed in social interaction, defined as an answer to a *why* question [21,23,39].

Lewis [21] asserts that most explanations are typically answers to *why* questions, for example, "Why did you do that?" or "Why did that happen?" *Why* questions are implicitly contrastive, with the form of the question being "why that decision *rather than something else?* [21]." When humans answer a *why* question, we offer an explanation (P) relative to some other event that did not occur (Q). This is termed a contrastive explanation, where P is referred to as the *fact* (an event that occurred) and Q is the *foil* (an event that did not occur) [39].

The factual component of a *why* question can have many potential foils. Consider the question "Why did you watch Game of Thrones?" Its foils could include "rather than the news?" or "instead of going out?" The foil itself generates context for the explanation to be provided. In human interactions, foils are often not explicitly stated in the *why* question. Instead, humans infer the foil from the tone and context of the interaction [39]. Importantly, the cause explained is dependent on the questioner's interest implied through conversation. This further emphasizes the assertion by Miller [23] that explanations are social and conversational [22].

Explaining the contrast can be easier than explaining the fact itself because P does not need to be sufficient for the event to occur, provided it differentiates between the causal difference of P and Q [40]. Contrastive explanations also have the benefit of constraining the information to be provided [39]. The constraining effect of a contrastive explanation is helpful to humans because it reduces the cognitive burden of processing an explanation [26].

Humans rarely provide an exhaustive causal chain as an explanation, preferring instead to select one or two pertinent causes [25,36]. In the example of the television show choice, a layperson would not justify the selection by providing their life story, listing influential childhood events. Instead, the explainer might answer, "It's more exciting than real life," and "I was tired," respectively. These explanations may not entirely explain P; however, they sufficiently and succinctly differentiate between P and Q. For the remainder of this paper, for simplicity, contrastive explanations will be referred to as *why*-explanations.

Explanation Complexity

Striking an appropriate balance in terms of the complexity of an explanation is difficult. Thagard's [41] theory of explanatory coherence states that people prefer simple, general explanations. This preference has been validated empirically [23]. For example, Read and Marcus-Newhall [42] evaluated this using a scenario about a woman with three symptoms: weight gain, fatigue, and nausea. The study participants received one of the following three explanation types describing what was causing her ill health:

1. Narrow:
   a. Having stopped exercising (explains weight gain).
   b. Having mononucleosis (explains fatigue).
   c. Having a stomach virus (explains nausea).
2. Broad: she is pregnant (explaining all three).
3. Conjunctive: all of the causes in 1 are true.

The participants preferred the broad explanation of pregnancy (option 2), preferring simple explanations that had fewer causes and explained more events.

Contemporary laboratory-based studies reiterate human preference for simple explanations over complex ones [13,36,43-45]. However, experiments in natural settings have revealed that complexity increases explanatory satisfaction [28,46] and that complexity preferences are aligned with the complexity of the event itself [47-49].

Frequently, studies in cognitive psychology examine diagnostic explanation through discussion of an alien race's illnesses to avoid reliance on prior knowledge. This removes a confounder; however, we should be cautious applying the findings to real-life SCs, given the relationship between explanation and knowledge (see *Explanation Seeking and Knowledge* section). Importantly, most of the experiments required high levels of literacy and comprehension. They were performed on cognitive psychology students or recruits screened for literacy ability. This is in contrast with reality: 1 in 7 of the UK population are functionally illiterate and would struggle to read a medicine label [50], and only half have an undergraduate degree [51].

In short, variance in explanatory preference has been noted between laboratory and natural experimental settings. In addition, experiments have been conducted on study groups with different characteristics to the general population. This reveals a need to validate preferences before generalizing to technologies used by the layperson population.

Explanations in AI

Overview

It should not be assumed that the presentation of explanations in AI systems matches with human explanatory behaviors and needs. Whereas the need for humans to explain themselves is often taken for granted in particular situations, there is continued debate around whether it is necessary for AIs to explain themselves. Turing Award winner Geoffrey Hinton argues that explanations are not necessary because humans cannot explain their own neural processes [52]. A study of medical students supports this because half of them were found to rely on intuitive thinking in diagnostic decision-making [53].

Recognizing that AI systems are frequently used to (help) make impactful decisions, explanations can be said to be required for at least two reasons. First, explanations are necessary for users to adopt AI technologies. When explanations are provided, trust and propensity to rely on systems increases [23,54-57]. If AI systems are not trusted, they are less likely to be adopted by their users, limiting the efficacy of the technology [19,58]. Second, the General Data Protection Regulation, the European Union law for data protection and privacy, places two relevant requirements on AI systems: they (1) must provide "meaningful information about the logic involved in the decision-making process" and (2) should ideally provide "an explanation of the decision [59]."

This results in a focus on explaining the decision-making of AI systems [57]. To do this, the algorithmic processes are examined to generate an explanation (the product), and the contained knowledge is subsequently communicated to the recipient (a social process) [23]. In contrast to human explanatory preferences, current AI explanations predominantly provide answers to *how* questions, not *why* questions.

How-Explanations in AI

AI models are often treated as black boxes due to their complexity and opacity. The explainable AI field predominantly focuses on increasing the transparency of how models produce their outputs, typically for use by computer programmers and expert users. This essentially answers a *how* question: "How did you decide that?"

There are a vast number of techniques that are used to explain how an AI model comes to a decision or otherwise produces an output [13]. These functional explanations will be referred to as *how*-explanations. Two such *how*-explanations are currently used to explain the outputs of SCs (see *Explanations in Symptom Checkers* section):

1. Input influence: Presents a list of the variables input to the model with a quantitative measure of their contribution (positive or negative) to the outcome [60]. Consider a system that reads mammograms. An input influence explanation might highlight visual areas of the scan that strongly influenced its diagnosis of a tumor.
2. Case-based reasoning: Displays a case from the model's training data that is closest to the one being classified [61]. In the mammogram example, a case-based reasoning explanation might declare the scan to be negative and provide another individual's mammogram to explain the verdict.

Both these explanatory methods require their human recipient to have domain knowledge to evaluate the explanation. Neither provides an explicitly contrastive explanation or an answer to a *why* question. Instead, a cognitive burden is placed on a human expert recipient to evaluate and contrast a large number of data points to determine whether they agree with the decision. In the mammogram example, a person not trained in radiology would have insufficient knowledge to understand either explanation, potentially resulting in the recipient perceiving them as poor quality.

When AI systems are deployed to wider society, system builders (such as software engineers and designers) must package the technology into software. For general audiences, the system builders take these *how*-explanations and translate them into a form comprehensible to a nonexpert [15]. Although this approach seems natural to many computer scientists and lawyers, it is not necessarily the best approach for SCs.

SCs, like many AI systems, are often presented as conversational agents or assistants. Humans are prone to anthropomorphism [62], and virtual assistants have many features that lead users to infer human-like agency in the assistants' behavior [63]. In human conversation, we prefer social *why*-explanations (see *Everyday Explanations* section). If SCs are to act as conversational agents, this suggests that the explanations being offered should adhere to norms of human conversation by creating a shared understanding of the decision made between the system and a human recipient [64]. Given that it is system builders who currently generate explanations for AI systems, the large body of work on explanations rooted in sociological, philosophical, and cognitive theories is underused [38].

Why-Explanations in AI

The conflict between layperson perceptions of *how*-explanations and *why*-explanations is the focus of this study. Given human preferences, the hypothesis is that there will be a preference for *why*-explanations of SC outputs. As such, a type of

why-explanation that has been proposed as more effective and accessible will be included, the counterfactual explanation.

Counterfactual explanations present how the factors considered to reach a decision must change for an alternative decision to be made [26,40,65]. A counterfactual explanation implicitly answers the *why* question "Why did you decide outcome P rather than outcome Q?" by examining what would happen if the variables $V = (v_1, v_2,...)$ were different [65,66]. Returning to the mammogram example, a user may ask "*Why* did you diagnose a tumor [rather than no tumor]?" to which the counterfactual explanation may state "If *these* pixels were not white, I would not have diagnosed a tumor."

Counterfactual explanations appear naturally in human cognition. They are a pattern that feature prominently in our day-to-day thoughts [67], with the capability to think counterfactually emerging around the age of two [68]. They are also contrastive, aligning with human explanatory preferences (see *Everyday Explanations* section). Counterfactual explanations are considered an efficient way of communicating causal reasoning [66,69]. They are effective in highlighting model discrimination [70] and, importantly for this work, they offer a type of *why*-explanation [27,40,66]. Counterfactual explanations do require cognitive effort on the recipient's part. However, by their nature, they bound the scope of explanation, reducing cognitive burden [69], and are advocated as a more accessible method for nontechnical users [65].

This study aims to investigate layperson trust in SC explanations. Humans have strong preferences for *why*-explanations, whereas technical methods developed to explain AI systems to date typically give *how*-explanations. Three explanation types of relevance have been identified for evaluation. To directly address this study's focus, the current state of explanations in SCs were examined. SCs, like other AI systems, currently tend to provide *how*-explanations.

Explanations in Symptom Checkers

To explore the current use of explanations in SCs, we surveyed 10 commercially available SCs (Multimedia Appendix 1). All SCs surveyed were presented in chatbot format to provide a natural mechanism for data collection from the user, mimicking their experience of speaking to a clinician. This further reinforces the view that "causal explanation takes the form of conversation and is thus subject to the rules of explanation [22]," that is, the recipient foremost requires a *why*-explanation.

Current SCs do not allow users to indicate what kind of explanation they require. Instead, the explanation type and content are predefined by system builders. The explanations provided are succinct, typically consisting of a single sentence. This succinctness is likely driven by software user-experience principles, which classify complexity as a detractor to technology adoption [71]. Again, this presentation contrasts with the rich explanations typically generated in the explainable AI field for expert users.

The SCs typically presented a form of explanation alongside the suggested diseases causing symptoms. They were observed taking two forms: (1) a contraction of an input influence explanation which provided one or two health symptoms that most positively influenced the SC's decision; (2) a social proof.

Social proofs, popularized to system builders by Cialdini [72] and Eyal [73] are based on extensive psychological studies that demonstrate that humans are both consciously and unconsciously susceptible to others' cues when making decisions. Social proof tactics can include cues such as user reviews and likes on social media. In the case of SCs, social proof is offered by explaining how many people with the same symptoms have previously been diagnosed with a particular disease. Generating a more detailed view of a social proof would involve providing details of other cases classified by the model, that is, a case-based explanation. Consequently, a social proof explanation can be viewed as a contraction of a case-based reasoning explanation.

This evaluation of current SCs shows that they either provide an input influence explanation or a social proof explanation, both of which are *how*-explanations.

Against this backdrop of the purpose of explanations, human explanatory preferences, and explanation use in SCs, this exploratory study seeks to answer 2 research questions:

Research question 1: Does the type of explanation affect a layperson's trust in the explanation provided by a symptom checker?

Research question 2: Is the person's level of trust affected by existing knowledge of a disease?

Methods

Experimental Design

To answer these questions, a 2×4 between-subjects experimental design was constructed, with the participants randomly assigned to a treatment group.

To answer research question 1, four explanation types were selected that reflect the state of the art in explanations of AI and SC outputs: input influence, social proof, counterfactual explanation, and no explanation. In the case of the no explanation type, no specific statement was presented that alluded to how or why the model came to a decision; it was included to provide a baseline for the perception of explanatory quality in chatbot interactions. In the mammogram example, the no explanation type would simply output, "This scan is negative."

To address research question 2, two diseases were selected that had differing levels of population awareness:

1. Migraine: This was chosen as the well-known disease because it affects 1 in 7 of the population [74], making it a disease with high population awareness whose symptoms are widely understood.
2. Temporal arteritis: This was selected as the lesser-known disease because it has a low incidence rate of approximately 0.035% in individuals aged above 50 years [75].

Both diseases involve head pain, which was selected for relatability as the majority of the population have experienced headaches [76]. Epidemiological data were used as a proxy for

layperson knowledge to limit study scope because knowledge, like trust, is an intangible variable to measure. The explanations shown are presented in Table 1. Analysis of variance (ANOVA) was planned to compare mean scores across the treatment groups. Given the limited prior research, the anticipated effect size was unknown, and a small effect size (Cohen's d=0.2) was chosen. G*Power (Heinrich Heine University) was used to calculate the sample size required for this effect size among 4 groups with a power β=.8. Erring on the side of caution, a goal of 200 participants per group was set.

Table 1. Explanations shown to participants for each combination of the four explanation types and two diseases.

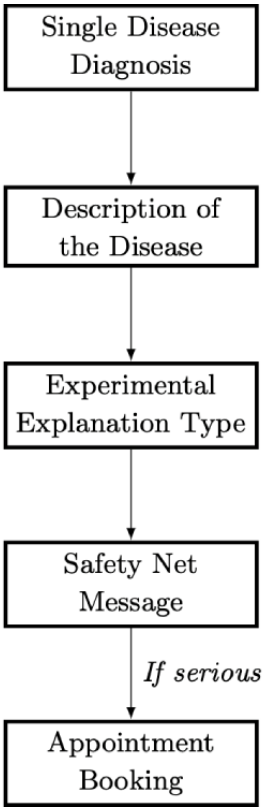
Explanation type and Disease	Explanation
Input influence	
Migraine	"I think this because you have a headache that came on slowly and you feel nauseous."
Temporal arteritis	"I think this because you have severe head pain in the side of your temples and your scalp and jaw are painful."
Social proof	
Migraine	"I think this because 8,217 people with your symptoms were diagnosed with a Migraine."
Temporal arteritis	"I think this because 8,217 people with your symptoms were diagnosed with Temporal Arteritis."
Counterfactual explanation	
Migraine	"If you didn't feel sick I'd have suggested you have a tension headache."
Temporal arteritis	"If your scalp and jaw didn't hurt, I'd have suggested you had a Migraine."
No explanation	
Migraine	No statement presented
Temporal arteritis	No statement presented

Stimulus Design

Chatbot simulation videos were created using a design and prototyping tool (Botsociety Inc). Information was presented in the manner of the Calgary-Cambridge model of physician-patient consultation [77], in line with the conversational presentation style used in current SCs. To avoid visual elements from confounding the experiment, the chatbot design was limited to text interaction. This contrasts with modern SCs, which use graphics to indicate additional explanatory factors such as the AI-model confidence.

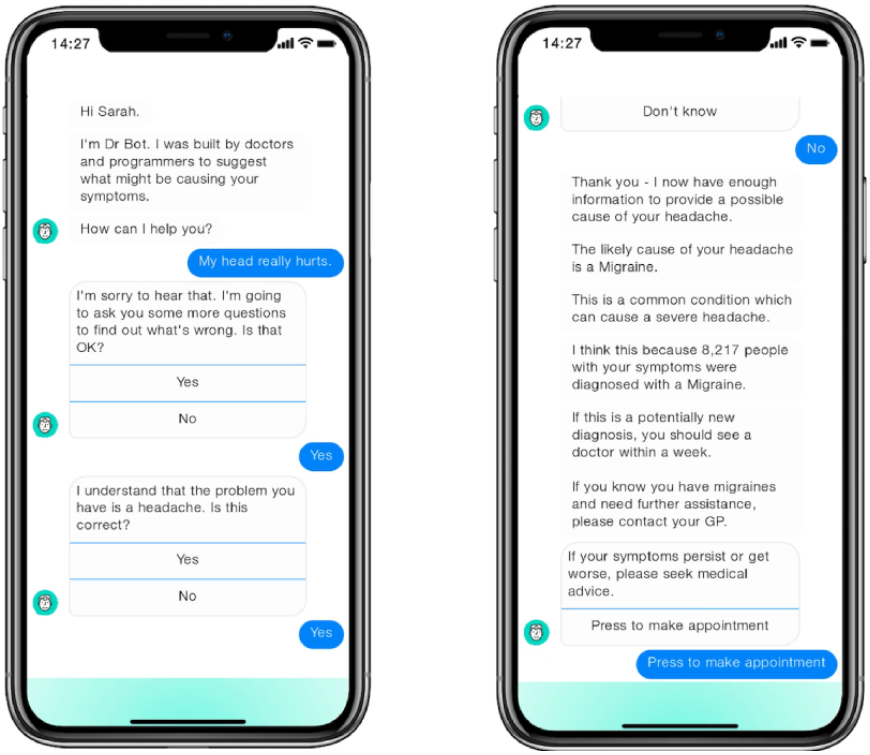
To verify stimuli presentation and question phrasing (see *Creating a Trust Measurement Scale* section), the stimuli and survey were piloted by conducting cognitive interviews with 11 individuals ranging in age from 28 to 62 years. The interviews revealed that at the end of the SC interaction, the participants expected to receive information that matched their typical consultation experiences. This included information about the disease and medical safety nets. Safety nets are a clinical management strategy to ensure the monitoring of patients who are symptomatic, with the aim of avoiding a misdiagnosis or nontreatment of a serious disease, for example, "If your symptoms persist or get worse, please seek medical advice." Consequently, these were included. The final SC videos were approximately 3 minutes in length, with most of the content being identical for each disease. Once the SC came to a conclusion, information was presented in the order shown in Figure 1.

Figure 1. The order of information presented at the end of the symptom checker flow.



The content of the explanations provided was set according to a medical advisor's clinical knowledge. For input influence, the two symptoms most likely to indicate the disease to a doctor were selected. Likewise, with counterfactual explanation, the symptom most likely to change a clinician's resulting opinion was chosen. For social proof, it was decided not to mirror current SC presentations (eg, "8/10 people with X had symptom Y") because it is known that probability information affects perceptions of explanations [36,44]. Instead, it was opted to state a large number of reference cases, as cognitive interviews indicated that lower numbers suggested a less-advanced AI model. Example screenshots of the stimuli are presented in Figure 2.

Figure 2. Screenshots of chatbot stimulus showing the start (left) and end (right) of social proof treatment of migraine.



Creating a Trust Measurement Scale

Trust is a hypothetical construct that cannot be observed or measured in a direct, physical sense [78]. Measurement scales for AI systems must minimally investigate two questions: "Do you trust the output?" (faith) and "Would you follow the system's advice?" (reliance) [79]. Experiments that have analyzed the reliability of trust scales have found high Cronbach's α values, suggesting that these scales are a reliable tool for measuring trust [80]. Nevertheless, there is a lack of agreement in the human-computer interaction field of how to measure trust in explanations. Hoffman [18] highlights that many scales are specific to the application context. Existing scales are also oriented toward evaluating opinions from expert users who repeatedly use a system over periods of time. This is in contrast to using an SC where an individual is likely to use it infrequently, that is, when they have a medical need.

This study's trust scale was based on Hoffman's Explanation Satisfaction Scale, with the scale tailored toward a layperson user [18]. Four categories of measurement were developed: faith (in the system), reliance (propensity to perform an action based on the explanation), satisfaction (attitude toward the explanation), and comprehension (of the explanation). In all, 3 questions were asked per category. The questions were inspired by scales developed by the human-computer interaction community [80-83], which had many commonalities. The full set of survey questions are presented in Tables S1 and S2 of Multimedia Appendix 2.

Data Collection

Participant Recruitment

Participants were recruited using the web-based platform Prolific. Qualtrics (Qualtrics, LLC) was subsequently used to randomly allocate the participants to a treatment group and gather survey responses. Participants who took less than 210 seconds to complete the experiment were removed as this would imply that they took less than 30 seconds to complete the survey, indicating satisficing. The final data set comprised 750 participants who were paid £1.45 (US $1.96 minimum wage equivalent) for their time.

Ethical Considerations

A number of ethical issues were considered in designing the study, including the following:

- Tests on individuals who are symptomatic require appropriate clinical trial procedures to be followed.
- If participants are not supervised by a clinician, an SC that is 100% accurate must be provided to avoid a misdiagnosis. Such an SC does not currently exist [5].
- Commercial SCs are designed to diagnose numerous diseases, resulting in a plethora of diagnostic pathways. SC model mechanisms are often nondeterministic, making real SCs uncontrollable in an experimental setting.
- The stimuli shown to participants must be medically safe and accurate to avoid misleading them.
- The act of observation could cause the development of health anxiety [84].

To mitigate each of these concerns, nonsymptomatic participants were asked to watch a video of an SC interaction. They were informed that this was not tailored to their own health needs. A General Medical Council–registered primary care physician experienced in SC design advised on the presented materials to ensure medical accuracy, safety, and realism. The participants were limited to individuals residing in the United Kingdom, which offers universal health care, and they were provided with resources to access clinical support if they became concerned. The study received ethical approval (reference: SSHOIIC1A19007) from the University of Oxford.

Results

Participants

Data from the 750 participants were analyzed using R (The R Foundation). The participants' ages ranged from 18 to 87 years (mean 35.8, SD 12.6). Of the 750 participants, 512 (68.3%) were in full-time or part-time employment. Most (723/750, 96.5%) had experienced a headache. The majority of respondents in the migraine treatment group viewed their disease as benign compared with a minority of those assigned to temporal arteritis (Table 2).

Table 2. Perceived seriousness of disease by percentage of respondents (N=750; migraine, n=367; temporal arteritis, n=383).

Seriousness	Disease	
	Migraine, n (%)	Temporal arteritis, n (%)
Not very serious	256 (69.8)	40 (10.4)
Moderately serious	105 (28.6)	266 (69.5)
Very serious	6 (1.6)	77 (20.1)

Exploratory Factor Analysis

The data were subsetted by topic, and exploratory factor analysis was performed to generate dependent variables from the 12 survey questions that measured trust. To test the impact of explanation type on layperson trust (RQ 1), the data were subsetted by disease, allowing the assessment of the impact of varying explanation types while holding disease constant. Similarly, to assess whether knowledge of a disease affects trust in explanation (RQ 2), the data were subsetted by explanation type, allowing the assessment of varying diseases while holding explanation type constant. As the measurement scale quantified different aspects of trust, the underlying factors could be correlated. To allow for correlation, oblimin (oblique) rotation was used.

Depending on the subset, two or three dependent variables emerged. In cases where the loadings resulted in two variables,

these were consistently interpreted as attitudes of *Faith* and *Comprehension*. In cases where the loadings produced three variables, themes of *Faith* and Comprehension still emerged, alongside an additional *Depth* variable. In this context, *Faith* was defined as blind trust in the explanation itself, *Comprehension* as an understanding of the information provided, and *Depth* as the richness of information provided.

For loadings of trust by disease, see Table 3, and for loadings of trust by explanation type, see Table 4.

Table 3. Summary of items, factor loadings, and correlations by disease with oblimin rotation where factors are displayed with their interpreted variable names (N=750; migraine, n=367; temporal arteritis, n=383)[a].

Question	Temporal arteritis		Migraine		
	Faith	Comprehension	Faith	Comprehension	Depth
16[b]	0.83	N/A[c]	0.80	N/A	N/A
17	0.83	N/A	0.79	N/A	N/A
18	0.74	N/A	0.68	N/A	N/A
19	0.58	N/A	0.62	N/A	N/A
20	0.54	N/A	0.57	N/A	N/A
21	0.78	N/A	0.77	N/A	N/A
22	0.75	N/A	0.74	N/A	N/A
23	0.55	N/A	N/A	N/A	0.57
24	__[d]	__[d]	N/A	N/A	0.92
25	N/A	0.88	N/A	0.83	N/A
26	N/A	0.71	N/A	0.65	N/A
27	N/A	0.46	N/A	0.42	N/A
28	N/A	0.79	N/A	0.76	N/A
Factor correlations					
Faith	__[e]	N/A	__[e]	N/A	N/A
Comprehension	0.58	__[e]	0.50	__[e]	N/A
Depth	__[d]	__[d]	0.52	0.39	__[e]

[a]For clarity, only factor loadings>0.4 are presented.

[b]Survey questions are presented in Table S1 of Multimedia Appendix 2.

[c]N/A: not applicable.

[d]Loadings of questions that were removed and correlations of factors not generated by analysis.

[e]Self-correlations.

Table 4. Summary of items, factor loadings[a], and correlations by explanation type with oblimin rotation where factors are displayed with their interpreted variable names (N=750; input influence, n=189; social proof, n=183; counterfactual explanation, n=192; no explanation, n=186).

>Question	Input influence		Social proof			Counterfactual explanation		No explanation		
	F[b]	C[c]	F	C	D[d]	F	C	F	C	D
16	0.88	N/A[e]	0.77	N/A	N/A	0.81	N/A	0.76	N/A	N/A
17	0.75	N/A	0.65	N/A	N/A	0.86	N/A	0.82	N/A	N/A
18	0.63	N/A	0.71	N/A	N/A	0.68	N/A	0.68	N/A	N/A
19	0.64	N/A	0.59	N/A	N/A	0.46	N/A	0.60	N/A	N/A
20	0.44	N/A	0.65	N/A	N/A	0.56	N/A	0.66	N/A	N/A
21	0.79	N/A	0.80	N/A	N/A	0.82	N/A	0.78	N/A	N/A
22	0.71	N/A	0.70	N/A	N/A	0.71	N/A	0.83	N/A	N/A
23	__[f]	__[f]	N/A	N/A	0.65	0.43	N/A	N/A	N/A	0.66
24	0.43	N/A	N/A	N/A	0.71	0.46	N/A	N/A	N/A	0.90
25	N/A	0.90	N/A	0.75	N/A	N/A	0.85	N/A	0.86	N/A
26	N/A	0.69	N/A	0.60	N/A	N/A	0.73	N/A	0.70	N/A
27	N/A	0.50	0.41	N/A	N/A	N/A	0.44	N/A	0.59	N/A
28	N/A	0.78	N/A	0.79	N/A	N/A	0.74	N/A	0.86	N/A
Factor correlations										
F	__[g]	N/A	__[g]	N/A	N/A	__[g]	N/A	__[g]	N/A	N/A
C	0.54	__[g]	0.47	__[g]	N/A	0.51	__[g]	0.50	__[g]	N/A
D	__[f]	__[f]	0.47	0.35	__[g]	__[f]	__[f]	0.59	0.45	__[g]

[a]For clarity, only factor loadings >0.4 are presented.

[b]F: faith.

[c]C: comprehension.

[d]D: depth.

[e]N/A: not applicable.

[f]Loadings of questions that were removed and correlations of factors not generated by analysis.

[g]Self-correlations.

Comparison of Explanation Trust by Varying Explanation Type

Variances in trust sentiment by explanation type were examined by performing multivariate ANOVAs (MANOVAs) on each disease subset. The MANOVA tests were followed by separate ANOVAs performed on each dependent variable.

For temporal arteritis, MANOVA suggested a marginal effect of explanation type on explanation trust, V=0.0289; F_{6758}=1.85; P=.09. ANOVAs revealed nonsignificant treatment effects on *Faith*, F_{3379}=1.32; P=.27 and *Comprehension*, F_{3379}=2; P=.11.

For migraine, MANOVA revealed no significant effect of explanation type on explanation trust, V=0.0187; $F_{91,089}$=0.759; P=.65. ANOVAs revealed nonsignificant effects on *Faith*, F_{3363}=0.7; P=.55; *Comprehension*, F_{3363}=1.13; P=.34; and *Depth*, F_{3363}=1.34; P=.26.

Comparison of Explanation Trust by Varying Disease

To investigate varying the disease presented, MANOVAs were performed on each explanation type subset using a single independent variable: the disease provided. MANOVA tests were followed by two-tailed t tests performed on each dependent variable.

Input influence (P=.001), social proof (P=.049), and no explanation (P=.006) were found to be significant, with counterfactual explanation (P=.053); t tests were used as a post hoc test. Means and SDs are presented in Table 5, and the parametric test results are reported in Multimedia Appendix 3.

Table 5. Mean scores and SDs for trust in explanation by varying disease presented (N=750; input influence, n=189; social proof, n=183; counterfactual explanation, n=192; no explanation, n=186).

Explanation type and disease	Faith, mean (SD)	Comprehension, mean (SD)	Depth, mean (SD)
Input influence			
Migraine	–0.141 (1.12)	0.0799 (0.968)	N/A[a]
Temporal arteritis	0.137 (0.736)	–0.0774 (0.902)	N/A
Social proof			
Migraine	–0.116 (0.931)	0.0661 (0.883)	–0.0614 (0.939)
Temporal arteritis	0.113 (0.955)	–0.0639 (0.917)	0.0594 (0.826)
Counterfactual explanation			
Migraine	0.023 (0.856)	0.144 (0.839)	N/A
Temporal arteritis	–0.022 (1.042)	–0.142 (0.984)	N/A
No explanation			
Migraine	0.0381 (0.974)	0.219 (0.823)	0.0134 (0.913)
Temporal arteritis	–0.035 (0.946)	–0.201 (1)	–0.0123 (0.936)

[a]N/A: not applicable.

Inclination to Use the SC

An overview of situations where participants would consider using an SC of this nature is presented in Table 6. Chi-square tests revealed no significant differences in these responses. For brevity, statistics are not reported here.

Table 6. Inclination to use symptom checkers (N=750; respondents were permitted multiple answers).

When would you use this type of symptom checker?	Respondents, n (%)
I would never use this kind of symptom checker	64 (8.5)
Any time I felt poorly	126 (16.8)
If I felt moderately unwell	317 (42.3)
If I couldn't speak to a human clinician	328 (43.7)
In situations where I would currently Google my symptoms	383 (51.1)

Discussion

Overview

Relatively little is known about the effect of explanations on layperson users' trust in increasingly ubiquitous AI SCs. This study aims to investigate the impact on trust among laypeople of *how*-explanations, which are commonly used by system builders, alongside a theoretically grounded *why*-explanation.

The findings reveal nuanced effects of the explanation that are first examined through varying the explanation type (see *Impact of Varying Explanation Type* section) and subsequently through varying the disease (see *Impact of Varying Disease* section). Finally, the participants' high propensity to use SCs is discussed, indicating that suboptimal explanations may be constraining SCs' ability to deliver health care more effectively (see *Propensity to Use SCs* section).

Impact of Varying Explanation Type

Factor Analysis as an Indication of Trust and Knowledge

As exploratory factor analysis generated different components per disease, the participants seemed to have differing conceptualizations of trust. The clean loading of the migraine responses into three components suggests greater nuance in interpretation compared with the less clear factor structure of temporal arteritis. Temporal arteritis did not generate a *Depth* component, and factor analysis required the removal of a question related to *sufficient detail*. As indicated previously, temporal arteritis was selected with the expectation that recipients generally knew less about temporal arteritis than about migraine.

Given the link between explanation and understanding [37,85], it follows that a recipient's perception of explanation quality gives insight into their mental model of the explanandum. The lack of a clear perception of *Depth*, coupled with the unclear factor loadings, suggests that the participants had less knowledge of temporal arteritis. In other words, there was a larger information gap between participants' general medical knowledge of temporal arteritis compared with their knowledge of migraine. Where an information gap occurred, an asymmetric knowledge condition was created where the participants had a need for knowledge to be transmitted [63].

Impact of Explanation Type in Trust of Migraine Explanation

Varying the type of explanation provided for migraine resulted in no significant difference in explanatory trust. This is likely attributable to the participants' knowledge of migraine.

Studies have demonstrated that patients evaluate medical information against their underlying knowledge base [86-88]. It is likely that as the participants watched the symptoms being described in the experiment, they relied upon their existing knowledge and formed a hypothesis of what may be the cause. As migraine symptoms are commonly known, the subsequent diagnosis of migraine aligned with the participants' knowledge.

An information gap is strongly predictive of an individual's need to be provided with an explanation [34]. The migraine results are supported by evidence that explanations are only required when there is an information gap between the explanation and existing knowledge [33]. In addition, because migraine is a common disease, it is highly probable; it is known that high priors increase the acceptance of explanations [36]. The lack of effect suggests that for a common, benign, well-known disease, the participants generally did not ask *why* the SC chose the diagnosis or *how* it came to the answer. Either the diagnosis made sense, or it did not. Thus, the explanation is not evaluated [14], and there is no impact of the explanation on the users' trust.

Impact of Explanation Type on Trust of Temporal Arteritis Explanation

Varying the type of explanation provided for temporal arteritis was reported as marginally significant (P=.09). This claim is made because significant findings when examining the effect of varying the disease (see *Impact of Varying Disease* section) add further support to the evidence that an effect is being observed. It is likely that the need for an explanation is stronger for temporal arteritis than for migraine. This explanatory need caused the participants to critically evaluate the temporal arteritis explanation. Before exploring these needs, it is instructive to first consider the stimulus that the participants were shown.

The chatbot video ends by informing the user that the disease requires immediate clinical attention and showing the user booking an appointment with a general practitioner. This severity was noted by the participants, 90% (345/383) of whom judged temporal arteritis to be moderately or very severe (Table 2). When presented with a diagnosis of temporal arteritis, most of the participants would not be knowledgeable of its symptoms. A diagnosis of an unknown, severe disease may have surprised or even alarmed the participants, causing an emotional response.

There are two potential factors that stimulated a greater need for explanation in the temporal arteritis group. First, as a rare disease with lesser-known symptoms, temporal arteritis created an information gap that caused a desire for an explanation (see *Explanation Seeking and Knowledge* section). Second, human emotion is known to affect explanation [37] and can influence how we experience events [89]. Surprise is a known predictor of explanation seeking when evaluating how a stimulus aligns with prior beliefs [34,90]. It is therefore possible that emotions such as surprise or even fear generated by the diagnosis may have provoked a greater need for an explanation. As the measurement of emotion was beyond the scope of this study, this would be a promising direction for future research.

Returning to the main discussion regarding the participants' need for an explanation, the marginal results suggest that an explanation may have improved trust, but there was no clear effect. The lack of significance was surprising, given the presence of a no explanation experimental treatment. Considering the result in the context of the experimental format illuminates possibilities as to why this should be so.

All experimental treatments, even no explanation, involved viewing a 3-minute chatbot interaction. The conversational nature of the interaction transmitted explanatory information [22]. The participants would have consulted their existing medical knowledge to build a hypothesis of the diagnosis while viewing the video. Finally, the description of the disease itself contained factual information. Therefore, the participants likely viewed all treatments, including no explanation, as a form of explanation. This is despite the no explanation type providing no explicit detail of *how* or *why* the SC derived its diagnosis.

The three remaining explanation types provided an answer to a *why* or *how* question. When framed by the need to close the temporal arteritis information gap, it seems that these explanations did not fully succeed. It seems that the explanation was not complete enough to be persuasive [45]. Lack of completeness in a medical explanation highlights a serious issue. There are tens of thousands of illnesses and conditions that can affect humans [91]. Of these, a layperson likely knows a handful of common illnesses and those that are the subject of public awareness campaigns. To a layperson, the explanations presented likely did not rule out other diseases that they were aware of. For example, when a layperson evaluated the symptoms presented by the counterfactual explanation, might they have wondered if headache and jaw pain also indicated a stroke? Returning to facts and foils (see *Everyday Explanations* section), it is logical that individuals can only generate foils of diseases that they are aware of. This is important because currently SCs have no awareness of the specific foil each user generates.

It is possible that the SC explanations are viewed as less effective because of a phenomenon known as causal discounting: "The role of a given cause in producing a given effect is discounted if other plausible causes are present [92]." In this scenario, the cause (temporal arteritis) is discounted in producing the effects (health symptoms) because other diseases could plausibly cause them.

Further support for causal discounting can be found in the text comments left by the participants upon survey completion. Of the 750 participants, 195 (26%) left comments. Of these 195, 59 (30.3%) indicated that they were fearful that a more serious disease may have been missed, equally distributed between the two diseases. This is a common concern, with 55% of patients fearing misdiagnosis [93]. As such, layperson need for reassurance that other serious diseases have been considered and dismissed is an important factor when explaining diagnoses. Although these comments were equally distributed between the

two diseases, the need to address causal discounting was greater for the disease that created the larger information gap.

It is possible that the simple explanations provided in this study were not sufficiently complete to close the recipient's information gap [46] and answer their foil. Although humans may prefer simple explanations [44], they do not blindly prefer them, instead calibrating their preferences according to the context of the explanandum [47]. This study's findings support those in previous works where explanations that are viewed as incomplete are less satisfying to the recipient [30,48]. In the case of the complex nuances of a diagnosis, this study aligns with the findings that people expect a level of complexity that matches that of the event itself [48]. This is orthogonal to the conversational maxim by Grice [94] that explanations should not suggest causes they think the explainee knows. In an SC scenario, this study suggests that the SC must explain that it has considered diseases that the explainee is aware of.

Having evaluated the effect of varying the explanation type on trust in an explanation for a given disease, this discussion now moves to the second perspective of this study: holding the explanation type constant and varying the disease being explained.

Impact of Varying Disease

Overview

This section builds on the previous finding that the disease being explained is a key factor in determining trust in explanation. As the findings are nuanced, this section sequentially examines the results of each explanation type. Two general themes emerged from this line of enquiry: (1) the uniqueness of symptoms better closes information gaps, and (2) different explanation types prime users to respond according to the explanation's emphasis. In cases where the emphasis is at odds with the user's own mental model, this causes cognitive dissonance, highlighting the danger of providing explanations when a user's explicit foil is not understood.

Before entering this exploration, it should be noted that although the MANOVAs conducted on three of the four explanation types were significant, post hoc *t* tests were not necessarily so (Table S2 of Multimedia Appendix 2). This suggests that the facets of trust in an explanation combine to create effects. It is also acknowledged that by dividing the respondents into the four explanation type groups, the post hoc tests were underpowered, and larger sample sizes were required for validation.

Input Influence

The most interesting finding when examining results by explanation type are that with regard to input influence, *Faith* is significantly different between the diseases (P=.053), with the temporal arteritis explanation more trusted than the migraine explanation (Table 5). The explanation given for temporal arteritis states a headache with scalp and jaw pain whereas migraine gives slow onset headache and nausea (Table 1). Returning to the discussion of the simplicity-complexity paradox and causal discounting, the distinctiveness of the combined symptoms given for temporal arteritis likely better closed the participants' information gap [95]. These unusual symptoms provide a simple explanation that excludes other diseases that the explainee may be aware of, cultivating greater faith [46]. Meanwhile, headache and nausea are symptoms that could be explained by diseases other than migraine, rendering it less trustworthy. As *Faith* explains the largest amount of variance in the factors of trust, the significance here underscores the marginal result when examining trust by explanation types for temporal arteritis.

Viewing input influence through this lens raises another important point. It is impossible to tell whether the participants perceived the explanation to be a *how*-explanation or a *why*-explanation. The aforementioned reasoning indicates that the participants were making a contrastive evaluation (indicating *why*); yet, we know that an expansion of the input influence explanation would result in a *how*-explanation (see *How-Explanations in AI* section).

Social Proof

Social proof is the only explanation format, which is clearly answering *how* the SC generated a diagnosis. It is known that the provision of an explanation can influence the importance of different features in understanding category membership [96,97]. In addition, laypeople observing AI systems are known to construct an internal mental model of the cognitive process of the software itself [98,99]. Interestingly, question 27—"It's easy to follow what the system does."—loaded into *Faith*, whereas for other explanation types, this question loaded into *Comprehension* (Table 4). This suggests that the *how* cue of social proof changed the participants' perception of the SC's cognitive mechanism. The participants were implicitly trusting social proof as a clustering technique as opposed to understanding the mechanism itself.

It is surprising that social proof generated a *Depth* factor, given the nature of this explanation. Return to the conversational nature of the SC causing participants to form a mental model of the disease as they observe questions being answered. By not providing an explanation which contains medical information, the recipient is left to evaluate their mental model against the diagnosis. This raises the question: does providing some form of explanation create more questions than it answers for laypeople?

Counterfactual Explanation

Comprehension was found to be significant (P=.03), with a small effect size and migraine being better comprehended than temporal arteritis (Table 5). Again, this points to the previous debate around information gaps and existing medical knowledge. The counterfactual explanation for migraine *undid* nausea, a well-known migraine symptom, matching general knowledge of migraines. For temporal arteritis, suggesting "If your scalp and jaw didn't hurt, I'd have suggested you had a migraine" did not sufficiently close the information gap for the participants. In the example, the user *did* have these symptoms. Erasing their existence did not rule out other diseases that the participants were aware of. It is not inconceivable that the participants saw the counterfactual explanation and thought, "But they do have scalp and jaw pain, couldn't these be caused by another serious condition like a stroke?"

Although counterfactual explanations may be theoretically advocated by social science as being effective in SC-type situations, a good explanation must be "relevant to both the question and to the mental model of the explainee [23,95,100]." The counterfactual explanations provided in this study, although medically valid, did not address the participants' *why* questions. This led to a significant impact on comprehension between the diseases; yet, there was no significant impact when comparing explanation types for the same disease. Existing literature calls for the use of counterfactual explanations [38], typically oriented toward expert users. As information gaps are smaller for expert users than for laypeople, the level of information required for transmission may differ. For example, a medical physician would be capable of ruling out serious, common diseases when observing the SC video, whereas the study findings indicate that a layperson may not be able to do so. Again, this stresses the importance of understanding a layperson's specific foil before generating an explanation.

No Explanation

With regard to the no explanation type, *Comprehension* was significant (P=.002), with a medium effect size and migraine being comprehended better than temporal arteritis. Notably, the factors generated for no explanation explained approximately 10% more of the variance in data than the factors for other explanation types (Multimedia Appendix 2). The no explanation format did not explain *how* the SC worked or *why* it suggested a particular disease. Again, the participants were left to consult their own medical knowledge. These results build on the discussion that providing an explanation may provoke doubt in the recipient by misaligning with their own mental model of both medical knowledge and of how the SC derives results (see *Social Proof* section).

Propensity to Use SCs

Despite middling levels of trust in the SC explanations, 91.5% (686/750) of the participants stated that they would consider using this type of SC (Table 6). This demonstrates that a large proportion of the digitally literate lay population would be prepared to use an SC when in real health need.

Of the 367 participants in the migraine group, 211 (57.5%) agreed or strongly agreed that the patient should follow the medical advice provided, as did 71.2% (273/383) of those shown temporal arteritis. It seems that a large proportion of the participants felt that a different intervention was necessary. It is not possible to tell whether the participants who were shown temporal arteritis and felt that the patient should not see a general practitioner urgently believed that the symptoms were benign (and that the patient should stay at home) or so serious that an ambulance should be called. The participants' disagreement with the triage instruction could be due to a concern that other diseases that they believe possible had not been considered and discounted, requiring further medical investigation. It could also be that they perceived that there was a greater emergency. This highlights that SC diagnoses are not subject to automation bias [101], the blind belief in a computer instruction because it has been generated by an intelligent system. Instead, users are skeptical of the SC's results. Returning to the societal principle of using SCs to reduce the burden on health care systems, this study shows that SC explanations must be improved to avoid a second human clinical opinion at an inappropriate triage level, increasing the burden on the health system.

Limitations

As this study was conducted to mirror the experience of a current SC, there are many confounding factors that could have affected the results. The explanations were examined in isolation; however, in a real-life scenario, intrinsic inputs such as pain, concern, and cognitive impairment may change layperson preference. It is also possible that the sample size was insufficient to detect small effects between the explanation types, given other confounding factors such as trust in the technology itself.

Ultimately, these results suggest that the field is ripe for further exploration. Productive lines for future inquiry could include measuring desire for explanation by disease, assessing information gaps and explanation seeking behaviors pre- and post explanation, emotion engendered by diagnosis and subsequent reaction to explanation, the level of complexity preferred in an SC diagnostic explanation, and understanding of diseases that users are concerned about. Crucially, future research must seek to understand users' foils in all these scenarios.

Conclusions

Millions of people around the world today are being encouraged to use SCs as a first port of call when seeking nonemergency medical care. Despite the prevalence of SCs, no specific research has been conducted into the effectiveness of delivering AI explanations to laypeople in the SC setting. SCs today provide explanations that explain *how* the AI cognitively derived its decision, although social science literature suggests humans prefer *why*-explanations. High-quality explanations are necessary to engender trust in an AI system, which is particularly important for SCs because lack of trust would cause users to seek second opinions from human clinicians. This additional demand could overburden health care systems and fail to deliver the promoted benefits of using SCs.

Our results suggest that the disease being explained is a primary factor in determining trust in subsequent explanation (RQ 2). This supports the view that when laypeople are presented with realistic scenarios, they use prior knowledge [102]. The disease diagnosed may or may not produce an information gap, resulting in differing needs for explanations and knowledge transmission. Our results show that an SC must explain the explanandum, as well as demonstrate that it has considered diseases that the explainee is aware of, to sufficiently close their information gap. Hence, transmitting knowledge is not as simple as selecting the pertinent cognitive elements of the model. Prior studies on medical diagnosis have abstracted away from human scenarios to isolate explanatory effects [13,30]. This study aligns with the findings that everyday, human explanations are far more nuanced and complex than laboratory-based experiments have suggested [48].

This study also provides some evidence that varying the explanation type affects layperson trust of the explanation (RQ

1), although these results are nuanced. For the well-known disease, information gaps were not cultivated; therefore, varying the explanation type had no effect. For the lesser-known disease, varying the explanation type resulted in marginally significant differences in trust. When the explanation type was held constant and the disease varied, three of the four explanations resulted in significant MANOVAs, indicating that facets of trust interact to create an overall perception of trust. Crucially, post hoc *t* tests, where significant, revealed that for some explanation types, temporal arteritis explanations were more trusted, and for some, migraine explanations were more trusted. The explanation types likely primed the participants to respond to particular signals highlighted within the explanations, resulting in these differences. Despite this, the explanation type did affect the level of trust in the explanation itself. These findings highlight a particular challenge of this study's design: by not knowing a participant's specific foil, the generically constructed explanations could not communicate sufficient knowledge, hampering the evaluation of the *how*-explanation and *why*-explanation formats.

The core finding of this study highlights that in order to close a user's information gap, the AI explanation must be generated with an understanding of that user's unique foil. System builders must not presume to know what question a layperson is asking of the system. Although system builders today work to elucidate the mechanisms of the AI system in a simple format, it is more important to close the gap between a layperson's general medical knowledge and the disease diagnosed. Part of this process must communicate that other diseases the user is aware of have been considered. Thus, system builders need to progress beyond communicating a simple, generic explanation toward the ability to receive real-time foils from a user.

Acknowledgments

Special thanks to Dr Keith Grimes, who consulted on the chatbot stimuli for medical validity and safety netting in his capacity as a National Health Service general practitioner. Thanks also to Dr Emily Liquin and Dr M Pacer for their stimulating conversations on explanations and mental models. This study would not have been possible without the support of the University of Oxford. Data collection was supported by grants from the Oxford Internet Institute and Kellogg College, Oxford. The article publication fee was provided by Open Access at the University of Oxford.

Conflicts of Interest

CW and DB were previously employed by, and hold share options in, Babylon Health, which has developed a symptom checker chatbot. Babylon Health had no input or involvement in this study. DB also declares current employment by, and shares in, Apple, which provides a COVID-19 symptom checker. Apple had no input or involvement in this study. BM currently serves in an advisory capacity for GSK Consumer Healthcare and has previously received reimbursement for conference-related travel from funding provided by DeepMind Technologies Limited.

Multimedia Appendix 1
Methodology and results of available Symptom Checker survey.
[DOCX File , 32 KB - jmir_v23i11e29386_app1.docx]

Multimedia Appendix 2
Survey questions asked to study participants.
[DOCX File , 21 KB - jmir_v23i11e29386_app2.docx]

Multimedia Appendix 3
Multivariate and univariate analyses of variance of trust in explanation by varying disease presented.
[DOCX File , 21 KB - jmir_v23i11e29386_app3.docx]

References

1. Universal Declaration of Human Rights. United Nations. URL: https://www.un.org/en/universal-declaration-human-rights/ [accessed 2021-09-23]
2. World report on ageing and health. World Health Organization. URL: https://www.who.int/ageing/publications/world-report-2015/en/ [accessed 2021-09-23]
3. Tracking universal health coverage: 2017 global monitoring report. World Health Organization and International Bank for Reconstruction and Development / The World Bank. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/259817/9789241513555-eng.pdf [accessed 2021-09-23]
4. Duke S. Babylon gets a healthy boost for its symptom-checking app. The Times. 2020. URL: https://www.thetimes.co.uk/article/babylon-gets-a-healthy-boost-for-its-symptom-checking-app-kfnwnqsxl [accessed 2021-09-23]
5. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]

6. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv. 2018. URL: https://arxiv.org/abs/1806.10698 [accessed 2021-09-23]
7. Buoy Health - Symptom Checker. Welcome. URL: https://www.welcome.ai/tech/deep-learning/buoy-health-symptom-checker [accessed 2021-09-23]
8. Haven D. An app a day keeps the doctor away. New Scientist. 2017. URL: https://www.newscientist.com/article/2141247-an-app-a-day-keeps-the-doctor-away/ [accessed 2021-09-23]
9. Hancock M. My vision for a more tech-driven NHS. Secretary of State for Health and Social Care Matt Hancock's speech at NHS Expo 2018. 2018. URL: https://www.gov.uk/government/speeches/my-vision-for-a-more-tech-driven-nhs [accessed 2021-09-23]
10. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
11. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
12. Ram A, Neville S. High-profile health app under scrutiny after doctors' complaints. The Financial Times. URL: https://www.ft.com/content/19dc6b7e-8529-11e8-96dd-fa565ec55929 [accessed 2021-09-23]
13. Narayanan M, Chen E, He J, Kim B, Gershman S, Doshi-Velez F. How do humans understand explanations from machine learning systems? An evaluation of the human-interpretability of explanation. arXiv.org. 2018. URL: http://arxiv.org/abs/1802.00682 [accessed 2021-09-23]
14. Herlocker J, Konstan J, Riedl J. Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work. 2000 Presented at: CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work; 2000; Philadelphia Pennsylvania USA. [doi: 10.1145/358916.358995]
15. Ribeiro M, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco California USA p. 1135-1144. [doi: 10.1145/2939672.2939778]
16. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
17. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
18. Hoffman R, Mueller S, Klein G, Litman J. Metrics for explainable AI: challenges and prospects. arXiv.org. 2018. URL: http://arxiv.org/abs/1812.04608 [accessed 2021-09-23]
19. Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. In: Proceedings of the 2015 International Conference on Healthcare Informatics. 2015 Presented at: 2015 International Conference on Healthcare Informatics; Oct 21-23, 2015; Dallas, TX, USA. [doi: 10.1109/ichi.2015.26]
20. Woodcock C. Code for reproducing results in "The Impact of Explanations on Layperson Trust in Artificial Intelligence–Driven Symptom Checker Apps: Experimental Study". GitHub. URL: https://github.com/clairewoodcock/impact-explanations-ai-sc [accessed 2021-09-23]
21. Lewis D. Causal explanation. In: Philosophical Papers Volume II. Oxford: Oxford Scholarship Online; 1987.
22. Hilton DJ. Conversational processes and causal explanation. Psychological Bulletin 1990;107(1):65-81. [doi: 10.1037/0033-2909.107.1.65]
23. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019 Feb;267:1-38. [doi: 10.1016/j.artint.2018.07.007]
24. Ruben D. Explaining Explanation. London, UK: Routledge; 1993.
25. Lombrozo T. The structure and function of explanations. Trends Cogn Sci 2006 Oct;10(10):464-470. [doi: 10.1016/j.tics.2006.08.004] [Medline: 16942895]
26. Lombrozo T. Explanation and abductive inference. In: The Oxford Handbook of Thinking and Reasoning. Oxford: Oxford Handbooks Online; 2012.
27. Hoffman R, Mueller S, Klein G. Explaining Explanation, Part 2: Empirical Foundations. IEEE Intelligent Systems 2017 Aug 17;32(4):78-86. [doi: 10.1109/mis.2017.3121544]
28. Herman B. The promise and peril of human evaluation for model interpretability. ArXiv. 2017. URL: http://arxiv.org/abs/1711.07414 [accessed 2021-09-23]
29. Meyer AN, Giardina TD, Spitzmueller C, Shahid U, Scott TM, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]

30. Liquin E, Lombrozo T. Inquiry, theory-formation, and the phenomenology of explanation. In: Proceedings of the 41st Annual Conference of the Cognitive Science Society. 2019 Presented at: 41st Annual Conference of the Cognitive Science Society; 2019; Montreal.
31. Loewenstein G. The psychology of curiosity: a review and reinterpretation. Psychol Bull 1994;116(1):75-98. [doi: 10.1037/0033-2909.116.1.75]
32. Wong W, Yudell Z. A normative account of the need for explanation. Synthese 2015 Feb 21;192(9):2863-2885. [doi: 10.1007/s11229-015-0690-8]
33. Mills CM, Sands KR, Rowles SP, Campbell IL. "I Want to Know More!": children are sensitive to explanation quality when exploring new information. Cogn Sci 2019 Jan;43(1):-. [doi: 10.1111/cogs.12706] [Medline: 30648794]
34. Liquin EG, Lombrozo T. A functional approach to explanation-seeking curiosity. Cogn Psychol 2020 Jun;119:101276 [FREE Full text] [doi: 10.1016/j.cogpsych.2020.101276] [Medline: 32062087]
35. Wilkenfeld DA, Plunkett D, Lombrozo T. Depth and deference: when and why we attribute understanding. Philos Stud 2015 May 7;173(2):373-393. [doi: 10.1007/s11098-015-0497-y]
36. Lombrozo T. Explanatory preferences shape learning and inference. Trends Cogn Sci 2016 Oct;20(10):748-759. [doi: 10.1016/j.tics.2016.08.001] [Medline: 27567318]
37. Keil FC. Explanation and understanding. Annu Rev Psychol 2006;57:227-254 [FREE Full text] [doi: 10.1146/annurev.psych.57.102904.190100] [Medline: 16318595]
38. Mittelstadt B, Russell C, Wachter S. Explaining explanations in AI. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Presented at: FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency; Jan 29 - 31, 2019; Atlanta GA USA. [doi: 10.1145/3287560.3287574]
39. Lipton P. Contrastive explanation. Roy Inst Philos Suppl 2010 Jan 08;27:247-266. [doi: 10.1017/s1358246100005130]
40. Lewis D. Causation. J Phil 1973 Oct 11;70(17):556-567. [doi: 10.2307/2025310]
41. Thagard P. Explanatory coherence. Behav Brain Sci 1989 Sep;12(3):435-467. [doi: 10.1017/s0140525x00057046]
42. Read SJ, Marcus-Newhall A. Explanatory coherence in social explanations: a parallel distributed processing account. J Personal Soc Psychol 1993 Sep;65(3):429-447. [doi: 10.1037/0022-3514.65.3.429]
43. Johnson S, Johnston A, Ave H. Explanatory Scope Informs Causal Strength Inferences. In: Proceedings of the 36th Annual Conference of the Cognitive Science Society. 2014 Presented at: 36th Annual Conference of the Cognitive Science Society; 2014; Quebec p. 2453-2458. [doi: 10.13140/2.1.3118.9447]
44. Lombrozo T. Simplicity and probability in causal explanation. Cogn Psychol 2007 Nov;55(3):232-257. [doi: 10.1016/j.cogpsych.2006.09.006] [Medline: 17097080]
45. Pacer M, Lombrozo T. Ockham's razor cuts to the root: simplicity in causal explanation. J Exp Psychol Gen 2017 Dec;146(12):1761-1780. [doi: 10.1037/xge0000318] [Medline: 29251989]
46. Kulesza T, Stumpf S, Burnett M, Yang S, Kwan I, Wong W. Too much, too little, or just right? Ways explanations impact end users' mental models. In: Proceedings of the 2013 IEEE Symposium on Visual Languages and Human Centric Computing. 2013 Presented at: 2013 IEEE Symposium on Visual Languages and Human Centric Computing; Sep 15-19, 2013; San Jose, CA, USA. [doi: 10.1109/vlhcc.2013.6645235]
47. Johnson SG, Valenti J, Keil FC. Simplicity and complexity preferences in causal explanation: an opponent heuristic account. Cogn Psychol 2019 Sep;113:101222. [doi: 10.1016/j.cogpsych.2019.05.004] [Medline: 31200208]
48. Lim JB, Oppenheimer DM. Explanatory preferences for complexity matching. PLoS One 2020 Apr 21;15(4):e0230929 [FREE Full text] [doi: 10.1371/journal.pone.0230929] [Medline: 32315308]
49. Vasilyeva N, Wilkenfeld D, Lombrozo T. Contextual utility affects the perceived quality of explanations. Psychon Bull Rev 2017 Oct;24(5):1436-1450. [doi: 10.3758/s13423-017-1275-y] [Medline: 28386858]
50. Leitch A. Leitch review of skills. GOV.UK. 2006. URL: https://www.gov.uk/government/organisations/leitch-review-of-skills [accessed 2021-09-23]
51. Higher Education Student Statistics: UK, 2017/18 - Qualifications achieved. HESA. 2019. URL: https://www.hesa.ac.uk/news/17-01-2019/sb252-higher-education-student-statistics/qualifications [accessed 2021-09-23]
52. Simonite T. Google's AI guru wants computers to think more like brains. Wired. 2018. URL: https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/ [accessed 2021-09-23]
53. Tay SW, Ryan PM, Ryan CA. Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. Can Med Ed J 2016 Oct 18;7(2):e97-103. [doi: 10.36834/cmej.36777]
54. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020 Jun;58:82-115. [doi: 10.1016/j.inffus.2019.12.012]
55. Hoffman RR, Klein G, Mueller ST. Explaining explanation for "Explainable Ai". In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2018 Presented at: Human Factors and Ergonomics Society Annual Meeting; 2018; Philadelphia p. 197-201. [doi: 10.1177/1541931218621047]
56. Clancey WJ. The epistemology of a rule-based expert system —a framework for explanation. Artif Intell 1983 May;20(3):215-251. [doi: 10.1016/0004-3702(83)90008-5]

57. Explaining decisions made with AI. Information Commissioner's Office. 2020. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/ [accessed 2021-09-23]
58. Zhu J, Liapis A, Risi S, Bidarra R, Youngblood G. Explainable AI for designers: a human-centered perspective on mixed-initiative co-creation. In: Proceedings of the 2018 IEEE Conference on Computational Intelligence and Games (CIG). 2018 Presented at: IEEE Conference on Computatonal Intelligence and Games, CIG; Aug 14-17, 2018; Maastricht, Netherlands. [doi: 10.1109/cig.2018.8490433]
59. Guide to the UK General Data Protection Regulation (UK GDPR). Information Commissioner's Office. URL: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/ [accessed 2021-09-23]
60. Datta A, Sen S, Zick Y. Algorithmic transparency via quantitative input influence: theory and experiments with learning systems. In: Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP). 2016 Presented at: 2016 IEEE Symposium on Security and Privacy (SP); May 22-26, 2016; San Jose, CA, USA. [doi: 10.1109/sp.2016.42]
61. McSherry D. Explaining the Pros and Cons of Conclusions in CBR. In: Proceedings of the 7th European Conference, ECCBR. 2004 Presented at: 7th European Conference, ECCBR; Aug 30- Sep 2, 2004; Madrid, Spain. [doi: 10.1007/b99702]
62. Epley N, Waytz A, Cacioppo JT. On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 2007;114(4):864-886. [doi: 10.1037/0033-295x.114.4.864]
63. Graaf MD, Malle B. How people explain action (and Autonomous Intelligent Systems Should Too). Explanations of behavior How people use theory of mind to explain robot behavior. 2017. URL: https://www.researchgate.net/publication/320930548_How_people_explain_action_and_AIS_should_too [accessed 2021-09-23]
64. Malle B. How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction. Cambridge: MIT Press; 2006.
65. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. arXiv.org. 2017. URL: https://arxiv.org/abs/1711.00399 [accessed 2021-09-23]
66. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. New York: Basic Books; 2018.
67. Gilovich T, Medvec VH. The temporal pattern to the experience of regret. J Pers Soc Psychol 1994;67(3):357-365. [doi: 10.1037/0022-3514.67.3.357]
68. Riggs K, Peterson D. Counterfactual thinking in pre-school children: mental state and causal inferences. In: Children's Reasoning and the Mind. UK: Taylor & Francis; 2000:87-99.
69. Byrne RM. Mental models and counterfactual thoughts about what might have been. Trends Cogn Sci 2002 Oct 01;6(10):426-431. [doi: 10.1016/s1364-6613(02)01974-5] [Medline: 12413576]
70. Kilbertus N, Rojas-Carulla M, Parascandolo G, Hardt M, Janzing D, Schölkopf B. Avoiding discrimination through causal reasoning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Long Beach California USA.
71. Krug S. Don’t Make Me Think, Revisited: A Common Sense Approach to Web Usability. 3rd edition. San Francisco: Peachpit; 2013.
72. Cialdini R. Influence: The Pyschology of Persuasian. New York: Harper Business; 2006.
73. Eyal N. Hooked: How to Build Habit-Forming Products. New York: Penguin; 2014.
74. GBD 2016 Headache Collaborators. Global, regional, and national burden of migraine and tension-type headache, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2018 Nov;17(11):954-976 [FREE Full text] [doi: 10.1016/S1474-4422(18)30322-3] [Medline: 30353868]
75. Watts RA. 2. Epidemiology of giant cell arteritis: a critical review. Rheumatology 2014 Jul 22;53(suppl 2):i1-i2. [doi: 10.1093/rheumatology/keu183]
76. Rasmussen BK, Jensen R, Schroll M, Olesen J. Epidemiology of headache in a general population--a prevalence study. J Clin Epidemiol 1991;44(11):1147-1157. [doi: 10.1016/0895-4356(91)90147-2] [Medline: 1941010]
77. Denness C. What are consultation models for? InnovAiT 2013 Sep 06;6(9):592-599. [doi: 10.1177/1755738013475436]
78. Muir BM. Trust in automation: part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 2007 May 31;37(11):1905-1922. [doi: 10.1080/00140139408964957]
79. Adams BD, Bruyn LE. Trust in Automated Systems. Department of National Defense. 2003. URL: https://cradpdf.drdc-rddc.gc.ca/PDFS/unc17/p520342.pdf [accessed 2021-09-19]
80. Jian J, Bisantz A, Drury C. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon 2000 Mar;4(1):53-71. [doi: 10.1207/S15327566IJCE0401_04]
81. Wang D, Yang Q, Abdul A, Lim B. Designing theory-driven user-centric explainable AI. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow, Scotland UK. [doi: 10.1145/3290605.3300831]
82. Madsen M, Gregor S. Measuring human-computer trust. In: Proceedings of the 11 th Australasian Conference on Information Systems. 2000 Presented at: 11 th Australasian Conference on Information Systems; 2000; Perth.
83. Cahour B, Forzy J. Does projection into use improve trust and exploration? An example with a cruise control system. Safety Sci 2009 Nov;47(9):1260-1270. [doi: 10.1016/j.ssci.2009.03.015]

84. Taylor S, Asmundson G. Exposure to other forms of disease-related information. In: Treating Health Anxiety: A Cognitive-Behavioural Approach. New York: Guilford Press; 2004:56.
85. Williams JJ, Lombrozo T. Explanation and prior knowledge interact to guide learning. Cogn Psychol 2013 Feb;66(1):55-84. [doi: 10.1016/j.cogpsych.2012.09.002] [Medline: 23099291]
86. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
87. Patel V, Groen G. Knowledge based solution strategies in medical reasoning. Cogn Sci 1986 Mar;10(1):91-116. [doi: 10.1016/S0364-0213(86)80010-6]
88. Scott SE, Walter FM, Webster A, Sutton S, Emery J. The model of pathways to treatment: conceptualization and integration with existing theory. Br J Health Psychol 2013 Feb;18(1):45-65. [doi: 10.1111/j.2044-8287.2012.02077.x] [Medline: 22536840]
89. Wilson TD, Centerbar DB, Kermer DA, Gilbert DT. The pleasures of uncertainty: prolonging positive moods in ways people do not anticipate. J Pers Soc Psychol 2005 Jan;88(1):5-21. [doi: 10.1037/0022-3514.88.1.5] [Medline: 15631571]
90. Frazier B, Gelman S, Wellman H. Preschoolers' search for explanatory information within adult-child conversation. Child Dev 2009;80(6):1592-1611 [FREE Full text] [doi: 10.1111/j.1467-8624.2009.01356.x] [Medline: 19930340]
91. International Statistical Classification of Diseases and Related Health Problems (ICD). World Health Organization. URL: https://www.who.int/classifications/classification-of-diseases [accessed 2021-09-23]
92. Kelley H. Causal schemata and the attribution process. In: Attribution: Perceiving the Causes of Behavior. New Jersey: Lawrence Erlbaum Associates, Inc; 1987:151-174.
93. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii21-ii27 [FREE Full text] [doi: 10.1136/bmjqs-2012-001615] [Medline: 23771902]
94. Grice H. Logic and conversation. In: Speech Acts. New York: Academic Press; 1975:41-58.
95. Hilton DJ. Mental models and causal explanation: judgements of probable cause and explanatory relevance. Think Reason 1996 Nov;2(4):273-308. [doi: 10.1080/135467896394447]
96. Giffin C, Wilkenfeld D, Lombrozo T. The explanatory effect of a label: explanations with named categories are more satisfying. Cognition 2017 Nov;168:357-369. [doi: 10.1016/j.cognition.2017.07.011] [Medline: 28800495]
97. Lombrozo T. Explanation and categorization: how "why?" informs "what?". Cognition 2009 Feb;110(2):248-253. [doi: 10.1016/j.cognition.2008.10.007] [Medline: 19095224]
98. Berzonsky MD. The role of familiarity in children's explanations of physical causality. Child Dev 1971 Sep;42(3):705-715. [doi: 10.2307/1127442]
99. Tullio J, Dey A, Chalecki J, Fogarty J. How it works: a field study of Non-technical users interacting with an intelligent system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI07: CHI Conference on Human Factors in Computing Systems; Apr 28 - May 3, 2007; San Jose California USA. [doi: 10.1145/1240624.1240630]
100. Byrne RM. The construction of explanations. In: AI and Cognitive Science '90. London: Springer; 1991.
101. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? Int J Hum Comput Stud 1999 Nov;51(5):991-1006. [doi: 10.1006/ijhc.1999.0252]
102. Zemla JC, Sloman S, Bechlivanidis C, Lagnado DA. Evaluating everyday explanations. Psychon Bull Rev 2017 Oct;24(5):1488-1500. [doi: 10.3758/s13423-017-1258-z] [Medline: 28275989]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
MANOVA: multivariate analysis of variance
RQ: research question
SC: symptom checker

Edited by R Kukafka; submitted 06.04.21; peer-reviewed by J Knitza, J Ropero; comments to author 28.06.21; revised version received 11.07.21; accepted 27.07.21; published 03.11.21.

Please cite as:
Woodcock C, Mittelstadt B, Busbridge D, Blank G
The Impact of Explanations on Layperson Trust in Artificial Intelligence–Driven Symptom Checker Apps: Experimental Study
J Med Internet Res 2021;23(11):e29386
URL: https://www.jmir.org/2021/11/e29386
doi:10.2196/29386
PMID:34730544

Review

Legal and Ethical Considerations for the Design and Use of Web Portals for Researchers, Clinicians, and Patients: Scoping Literature Review

Michael Lang[1], BA, BCL, LLB; Sébastien Lemieux[2], BSc, MSc, PhD; Josée Hébert[2,3,4], MD; Guy Sauvageau[2,4], MD, PhD; Ma'n H Zawati[1], LLB, LLM, DCL

[1]Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
[2]Department of Medicine, Université de Montréal, Montreal, QC, Canada
[3]The Leucegene Project at Institute for Research in Immunology and Cancer, Université de Montréal, Montreal, QC, Canada
[4]Division of Hematology, Maisonneuve-Rosemont Hospital, Montreal, QC, Canada

Corresponding Author:
Ma'n H Zawati, LLB, LLM, DCL
Centre of Genomics and Policy, Department of Human Genetics
Faculty of Medicine and Health Sciences
McGill University
740 Avenue Dr. Penfield
Montreal, QC, H3A 0G1
Canada
Phone: 1 (514) 398 8155
Email: man.zawati@mcgill.ca

Abstract

Background: This study aims to identify a novel potential use for web portals in health care and health research: their adoption for the purposes of rapidly sharing health research findings with clinicians, scientists, and patients. In the era of precision medicine and learning health systems, the translation of research findings into targeted therapies depends on the availability of big data and emerging research results. Web portals may work to promote the availability of novel research, working in tandem with traditional scientific publications and conference proceedings.

Objective: This study aims to assess the potential use of web portals, which facilitate the sharing of health research findings among researchers, clinicians, patients, and the public. It also summarizes the potential legal, ethical, and policy implications associated with such tools for public use and in the management of patient care for complex diseases.

Methods: This study broadly adopts the methods for scoping literature reviews outlined by Arskey and O'Malley in 2005. Raised by the integration of web portals into patient care for complex diseases, we systematically searched 3 databases, PubMed, Scopus, and WestLaw Next, for sources describing web portals for sharing health research findings among clinicians, researchers, and patients and their associated legal, ethical, and policy challenges. Of the 719 candidate source citations, 22 were retained for the review.

Results: We found varied and inconsistent treatment of web portals for sharing health research findings among clinicians, researchers, and patients. Although the literature supports the view that portals of this kind are potentially highly promising, they remain novel and are not yet widely adopted. We also found a wide range of discussions on the legal, ethical, and policy issues related to the use of web portals to share research data.

Conclusions: We identified 5 important legal and ethical challenges: privacy and confidentiality, patient health literacy, equity, training, and decision-making. We contend that each of these has meaningful implications for the increased integration of web portals into clinical care.

(J Med Internet Res 2021;23(11):e26450) doi:10.2196/26450

KEYWORDS

medical ethics; web portal; scoping review; eHealth; portal

Introduction

Background

Medical care and health research are jointly undergoing significant changes brought about by the internet [1-3]. New web portals, apps, and programs are helping to facilitate unprecedented levels of data sharing and collaboration, potentially enabling more precise targeted treatment and rapid research translation [4-6]. Web portals have been a significant part of this emerging web-based health ecosystem, providing patients with a mechanism for accessing electronic health records, managing appointments and prescriptions, and communicating directly with care providers [7]. Much has been written about the technical and ethical challenges associated with the development and integration of web portals into clinical settings [8,9]. However, portal technology might also be used to connect health researchers to clinicians, patients, and the public. Web portals, for example, could be a useful platform for broad and rapid dissemination of research results. Although the most prevalent and widely discussed types of web portals are those used to manage patient interactions with the health care system (so-called *patient portals*), other potential uses are being increasingly identified.

Objective

This study aims to identify one such novel potential use: the adoption of web portals to rapidly share health research findings with clinicians, scientists, and patients. In the era of precision medicine and learning health systems, the translation of research findings into targeted therapies depends on the availability of big data and emerging research results. Web portals may work to promote the availability of novel research, working in tandem with traditional scientific publications and conference proceedings. Web portals raise the possibility that important research findings will be shared efficiently not only with scientists and clinicians but also with patients and members of the public. This study aims to examine how web portals can be used to facilitate such sharing. We seek to preliminarily outline how web portals can be used to advance research sharing objectives and to identify legal and ethical barriers that might challenge these functions. We will distinguish between *web portals* as a general category, by which we mean web resources that aggregate health information for a specific purpose, and *therapeutic portals*, by which we mean web portals designed to aggregate health research findings and make them available to a diverse array of researchers, clinicians, and the general public. Therapeutic portals may be contrasted with one prominent type of web portal, the *patient portal*. This is a system typically used to share individual patient data, facilitate clinical interaction, and record specific health outcomes [10]. For the purposes of this study, we propose the concept of a *therapeutic portal* in an attempt to capture nuances that are not fully expressed in discussions on the use of web portals for managing and sharing health research findings. In the era of precision medicine and learning health systems, as interactions between clinicians, researchers, and the public take on a unique and pressing character, we are of the view that the language of therapeutic portals better fits the potential emerging use of portal technology in these contexts. Although the distinction between these kinds of web portals is subtle, it has potentially important implications for the provision of care and for the legal and ethical obligations of clinicians in health care systems increasingly modulated by the internet.

Our interest in studying the feasibility of therapeutic portals is motivated by an ongoing multidisciplinary molecular genetics research project to better characterize and categorize the incidences of acute myeloid leukemia (AML) [11]. The Leucegene project is a biobank-based study adopting next-generation sequencing, chemotherapeutics, and precision-medicine approaches for identifying prognostic markers and therapeutic targets for AML [12]. Research in this area, which brings together the efforts of researchers, clinicians who apply novel prognostic testing regimes, and patients who ultimately benefit from AML treatments derived from novel precision-medicine therapies, is fertile ground for exploring the use of internet-mediated tools for widely sharing advancements in the treatment and study of AML. Although using such systems promises to maximize the potential exploitation of new knowledge by making it accessible to as wide an audience as possible, it also raises difficult practical and ethical difficulties.

In this regard, this review aims to explore the potential use of web portals that work to share research results in the context of complex diseases with the research community, clinicians, and patients. In particular, we set out to understand whether such tools are presently in use, how they may be implemented, and how we might account for the pressing policy challenges they raise. This study will help contextualize and support the development of web portals intended to promote information sharing in health care systems by increasingly applying precision-medicine approaches to combat complex diseases. This review examines the potentially expanding scope and utility of web portals and highlights the policy challenges associated with this expansion. This will be achieved by outlining the results of a scoping literature review attempting to ascertain the extent to which web portals have been considered to facilitate communication between researchers and clinicians. Therapeutic portals are internet-enabled tools for sharing information between health researchers conducting basic science research and clinicians implementing best health practices. Tools of this kind expand on the increasing reach of patient portals by broadening their scope of application and inflating their role in patient care. Although there is a sense in which the use of web portals for sharing health research and disease data among clinicians and investigators is a logical extension of the form and function of existing tools that bring together patients and care providers, little direct consideration of such platforms appears to be present in the literature. This review poses 2 questions: (1) Are therapeutic portals being developed and used in health systems? (2) What legal and ethical issues might be raised by the use of therapeutic portals for sharing health research and disease data with researchers, clinicians, and the general public? This study will address these questions in 3 sections. First, we outline the methods used to preliminarily survey the literature in consideration of the questions above. Second, we summarize our results and describe how they apply to the development and use of therapeutic portals. Third, we discuss our findings with a broader literature on web portals

and health research dissemination. We also note the limitations of our research and suggest that more work on this topic is needed.

Methods

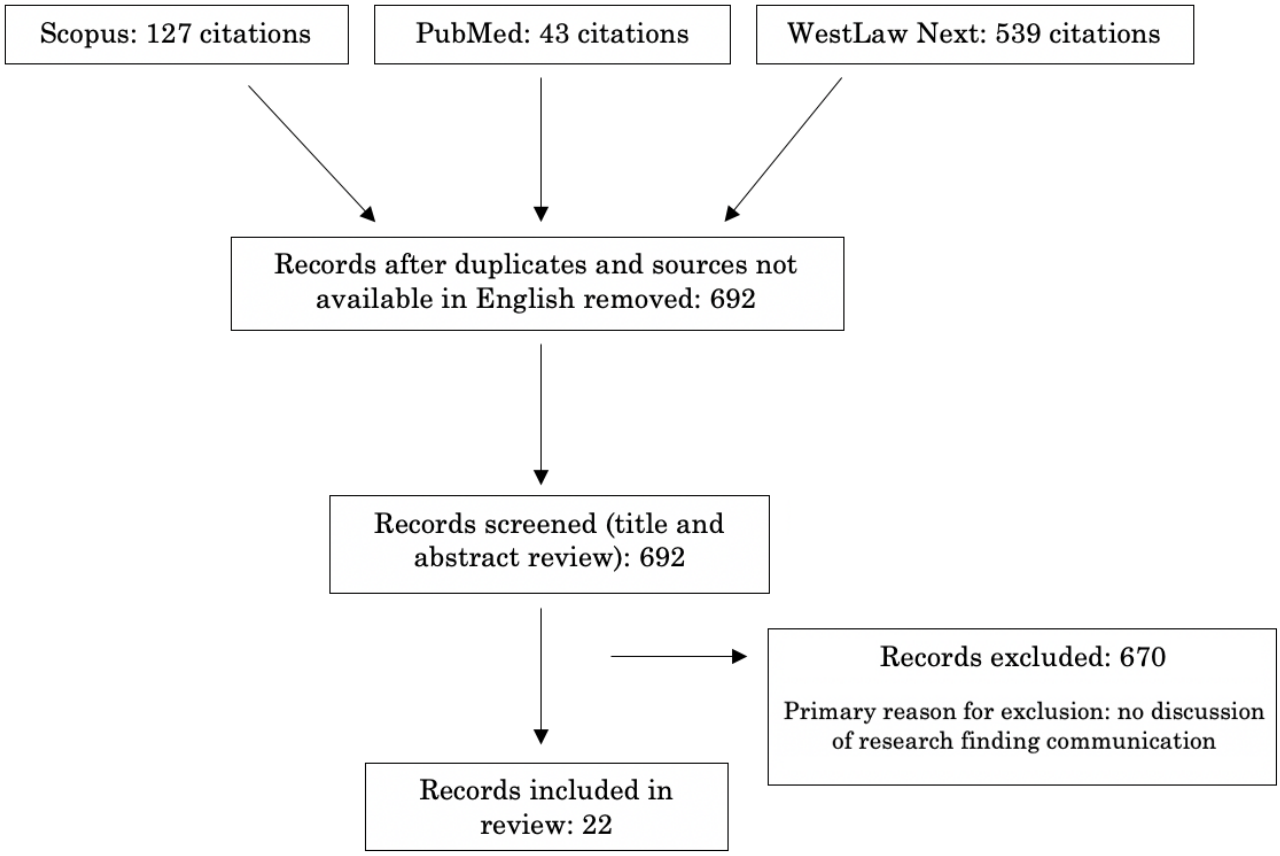
We designed this scoping literature review following the theoretical model set out by Arskey and O'Malley [13]. Scoping reviews are a social research methodology that aim to rapidly map *the key concepts underpinning a research area and the main sources and types of evidence available* [14]. Scoping reviews are typically contrasted with more extensive systematic reviews. This study, following Arskey and O'Malley, is intended to identify the broad contours of existing literature and note the gaps requiring further attention [13]. We adopted an iterative search process comprising 5 stages. First, we developed the following 2 research questions: (1) Are therapeutic portals being developed and used in health systems? (2) What legal and ethical issues might be raised by the use of therapeutic portals for sharing health research and disease data with researchers, clinicians, and the general public? Second, we identified relevant studies by searching web databases. Specifically, we consulted Scopus, PubMed/MEDLINE, and WestLaw Next using fixed Boolean keywords including *online portal* AND *health* AND *communication* AND *research** AND *clinic**. Our search strategy initially focused on web portals as a general category, noting that scholarly sources describing portals used primarily for communicating research findings to a wide audience were not organized using discretely identifiable terminologies. We refined search terms iteratively and scanned source references to broaden the scope of the review. A primary search was conducted in April 2019, and a final search was conducted in March 2020. We did not time limit our search to obtain the widest possible array of sources. Initially, we identified 719 articles and papers from each source. Third, we selected studies for inclusion according to the criteria developed *post hoc*, as we became increasingly familiar with the literature. Owing to the nature of the research questions, as well as inconsistent language in the literature for explaining the function of web portals, determining which sources to include in our analysis sample proved to require a delicate balance. We worked with a set of flexible inclusion criteria. We included sources that met any one of the following broad criteria: (1) description of a web portal in actual use for the purpose of sharing research findings or non–patient-specific data; (2) description of a potential use case for web portals that share research findings or non–patient-specific data among researchers, clinicians, and the public; (3) description of a potential use case for web portals that permit access to health research findings or lay explanations of health research findings for patients or the general public; and (4) descriptions of legal and ethical issues associated with any of the above criteria. We excluded (1) duplicate sources, (2) sources not available in English, and (3) sources that primarily discuss the use of web portals for purposes other than the sharing of health research. Exclusion criterion 3 resulted in a large number of sources being screened out. In particular, numerous sources focused narrowly on the development and implementation of *patient* portals were generally not included in the review. Of the 719 sources initially collected, 22 (3%) sources were retained for the final review. Fourth, we charted source data for each retained resource, including the author, year of publication, and conclusions. These data are outlined in Table 1. Fifth, we summarized and reported the results of the review in this study.

Results

Initial Screening

We selected 22 sources for the final review. Of the 719 sources initially obtained, 25 (3%) were eliminated as duplicates and 2 (0.2%) were unavailable in English. This left 96.2% (692/719) of papers for which we conducted a title and abstract review. Of these, we found that a high number primarily outlined work on the adoption of patient portals. Following exclusion criteria 3 above, those sources that did not include discussion about the potential communication of research findings, in addition to patient data such as testing results and treatment plans, were generally determined to fall outside the ambit of this review. We eliminated 93% (671/719) of sources that were determined not to meet the inclusion criteria described above. Operating on a relatively narrow conception of therapeutic portals, we did not expect to find a large number of sources that would meet the inclusion criteria outlined above. Nevertheless, we determined that the 3% (22/719) included sources provide an overview of this emerging field and suggest that the adoption of web portals for sharing research findings will become increasingly prominent in the coming years. This limited overview of the early literature in this space may work to ground subsequent discussions. Figure 1 outlines the steps taken to conduct this review.

Figure 1. Scoping review flowchart.



Study Selection

The sources retained for the final review were relatively varied in their theoretical and methodological orientations. Several studies, for example, described patient portals enhanced by the inclusion of health data that would not be included in patient charts. One such approach is exhibited in a work by Nordqvist et al [15], who outlined the attitudes of health professionals on a patient portal supplemented with disease-specific information intended to promote adolescent diabetes care. Other articles documented the creation of a novel web portal for safely sharing participant-level data with researchers [16,17] or an open web portal designed to facilitate the sharing of gene association data for lung cancer [18]. Other records included in our final sample mentioned issues surrounding research data sharing somewhat less directly [19]. Following the abstract and title review, we determined that these sources warranted inclusion in the final analysis. Where the precise contours of a proposed or existing portal were not clearly delineated, we erred on the side of the inclusion. This led to the analysis of a number of articles that do not directly contemplate what we have described as therapeutic portals, but that *do* contemplate use cases to facilitate access to health research findings. We believe that these sources fall under inclusion criteria 2, describing a potential use case for web portals that share research findings or non–patient-specific data among researchers, clinicians, and the public. We explore these themes in greater detail in the *Discussion* section.

Results Charting

We charted our overarching results in Table 1, which summarizes findings derived from sources reviewed at this final stage and conveys our determination on whether these sources describe tools that could reasonably be considered therapeutic portals. Of the 22 references, we established that 8 (36%) contemplated a web resource to facilitate some form of communication among researchers, patients or the public, and clinicians. The manner in which these resources describe therapeutic portals varies considerably.

Table 1. Methodological and theoretical orientations of reviewed articles.

Study and article title	Article objective	Article conclusions	Therapeutic portal
Baldwin et al [19], *Patient portals and health apps: Pitfalls, promises, and what one might learn from the other symptoms*	Limitations and potential of 2 kinds of patient-facing information technology: portals and applications	Combining features of mHealth apps and portals could increase patient engagement	No
Bartonova [20], *How can scientists bring research to use: The HENVINET experience*	Describes HENVINET, a portal for sharing research findings among scientists, policy makers, and the public	Portal highlights the need for liaison between researchers, policy makers, and the public	Yes
Bostrom et al [21], *Strategic and integrated planning for healthy, connected cities: Chattanooga case study*	Describes a web portal combining location and health data to identify areas for potential greenspace development	Portal determined access to park space. Promotes integrating community and social metrics to equitably address public health challenges	No
Bowler et al [22], *The visibility of health web portals for teens: A hyperlink analysis*	Assesses teen health websites for accessible and reliable health information	Websites had a low level of visibility relative to resources intended for other audiences. Information for teens present on resources that lack health expertise	Yes
Ling Cai et al [18], *LCE: An open web portal to explore gene expression and clinical associations in lung cancer*	Discusses a lung cancer database with expression and clinical data from 6700 patients in 56 studies	The Lung Cancer Explorer is publicly accessible and provides genomic and tissue image data for lung cancer	Yes
Christensen et al [23], *Beacon: A web portal to high-quality mental health websites for use by health professionals and the public*	Describes the Beacon web portal, which aggregates lists of high-quality health websites sharing information on mental health	There are many high-quality web resources available for mental health. The Beacon portal attempts to identify and gather them in a single resource	Yes
Das et al [24], *The impact of an eHealth portal on health care professionals' interaction with patients: qualitative study*	Discusses implementation of a web portal intended to help weight loss surgery patients achieve healthy outcomes	Implemented eHealth portal was a valuable source of information and a gateway for facilitating positive patient interactions	No
Kaiser [16], *A new portal for patient data*	Presents the Vivli portal, which is intended to support the sharing of anonymized clinical study data	The portal makes available the results of more than 4000 clinical trial data sets from 8 companies and nonprofits	No
Kirkpatrick et al [25], *GenomeConnect: matchmaking between patients, clinical laboratories and researchers to improve genomic knowledge*	Presents the GenomeConnect portal, which provides a space for patients and members of the public to share health history and genetic test results	GenomeConnect portal allows members of the public to participate in genetics research and to contribute to the validation of novel clinical tests	Yes
Kohler [26], *Can internet access growth help reduce the global burden of noncommunicable diseases?*	Describes an open access portal for linking disparate source health information for reducing preventable lifestyle-related risk factors associated with noncommunicable disease	Web portals of the kind described have the potential to improve the global burden of noncommunicable disease if implemented at scale	No
Kuijpers et al [27], *An interactive portal to empower cancer survivors: a qualitative study on user expectations*	Studies the perspectives of cancer survivors on the possible features of an interactive web portal	Participants interested in portal features that fulfill information needs, such as access to their eHealth record	No
Li et al [17], *Moving data sharing forward: the launch of the Vivli platform*	Presents the Vivli portal for supporting anonymized clinical study data sharing	Data sharing portals have an important role to play in addressing issues around reidentification and anonymization	No
Maggio et al [28], *Qualitative study of physicians' varied uses of biomedical research in the USA*	Perspectives of physicians on interaction with biomedical research presented on a web portal	Physicians reported a high level of research adoption and appealed to their multi-faceted roles as clinicians, educators, and researchers	No
Marrie et al [29], *Use of eHealth and mHealth technology by persons with multiple sclerosis*	Describes use of eHealth and mHealth systems by patients with multiple sclerosis	Internet-enabled tools help to facilitate the sharing of health information between clinicians and patients with multiple sclerosis	No
McKemmish et al [30], *Consumer empowerment through metadata-based information quality reporting: the breast cancer knowledge web portal*	Describes development of the BCK–Web portal, a web resource for sharing health information with patients with breast cancer	BCK–Web portal communicates high-quality medical and experiential knowledge	Yes

Study and article title	Article objective	Article conclusions	Therapeutic portal
Melchart et al [31], *Introduction of a web portal for an individual health management and observational health data sciences*	Establishes a set of core objectives and processes implementing a web portal for lifestyle changes and individual health management	Web tools help to facilitate individual health management in concert with health coaching	No
Melholt et al [32], *Cardiac patients' experiences with a telerehabilitation web portal: implications for eHealth literacy*	Explores use of a web portal cardiac patient rehabilitation. Outlines health literacy effects	A web portal for rehabilitation among cardiac patients may increase health literacy	No
Nordfeldt et al [33], *To use or not to use – practitioners' perceptions of an open web portal for young patients with diabetes*	Documents clinician perspectives on the use of an open access web portal for patients with juvenile diabetes	Clinicians felt comfortable recommending web resources for which available information was verifiably reliable	Yes
Nordqvist et al [15], *Health professionals' attitudes towards using a web 2.0 Portal for child and adolescent diabetes care: qualitative study*	Describes clinician perspectives on the use of a *Web 2.0* portal for juvenile patients with diabetes	Clinicians exhibited positive attitudes toward the portal. Support close collaboration between stakeholders in the development of future portals	Yes
Rocker [34], *Use of a web portal to facilitate clinical trial recruitment: preliminary analysis of fox trial finder*	Characterizes research volunteers registered on a web portal for clinical trial participation recruitment for the study of Parkinson's disease	Persons affected by Parkinson's disease willing to participate in health research and share personal data on the web	No
Sutherland et al [35], *A novel open access web portal for integrating mechanistic and toxicogenomic study results*	Describes collaborative toxicogeomics, a web portal for sharing *best practice* methods in computational biology	The developed open-source portal helps to increase accessibility, transparency, and collaboration between researchers in the field	No
Tomlinson et al [36], *MiMiR – an integrated platform for microarray data sharing, mining and analysis*	Presents MiMiR, a web portal supporting the management and sharing of microarray data	MiMiR portal contains more than 150 data points and over 3000 hybridizations supporting the microarray user community	No

Table 2 outlines the general thematic orientations of each of the 8 papers that we determined to have directly contemplated therapeutic portals. Beyond thematic diversity, we also found that reviewed sources considered a wide array of potential policy issues raised by the adoption of web portals in health care, both in the case of therapeutic portals and that of similarly disposed web portals. In particular, we found that 5 legal, ethical, and social issues were raised in the 22 resources we reviewed: (1) privacy and confidentiality, (2) patient health literacy, (3) equity, (4) training, and (5) decision-making. Overall, we determined that out of the 22 resources, 9 (41%) discussed privacy and confidentiality issues, 9 (41%) discussed patients' health literacy, 4 (18%) discussed equity, 9 (41%) discussed training, and 10 (45%) discussed decision-making. Table 3 outlines the distribution of these dominant policy issues among the reviewed sources.

Table 2. Thematic orientation of reviewed therapeutic portals (n=8).

Theme	Resources, n (%)	Study
Genomics	2 (25)	Cai et al [18]; Kirkpatrick et al [25]
Cancer	2 (25)	Cai et al [18]; McKemmish et al [30]
Diabetes	2 (25)	Nordfeldt et al [33]; Nordqvist et al [15]
Mental health	1 (12.5)	Christensen et al [23]
Teenage health	1 (12.5)	Bowler et al [22]
Health and the environment	1 (12.5)	Bartonova [20]

Table 3. Distribution of legal, ethical, and social issues (n=22).

Theme	Resources, n (%)	Citations
Privacy and confidentiality	9 (41)	Baldwin et al [19]; Bowler et al [22]; Das et al [24]; Kaiser [16]; Kirkpatrick et al [25]; Kuijpers et al [27]; Li et al [17]; McKemmish et al [30]; Melchart et al [31]
Patient health literacy	9 (41)	Baldwin et al [19]; Das et al [24]; Kirkpatrick et al [25]; Kohler [26]; Kuijpers et al [27]; McKemmish et al [30]; Melchart et al [31]; Melholt et al [31]; Nordqvist et al [15]
Equity	4 (18)	Bostrom et al [21]; Marrie et al [29]; McKemmish et al [30]; Melchart et al [31]
Training	9 (41)	Baldwin et al [19]; Bartonova [20]; Christensen et al [23]; Das et al [24]; Kuijpers et al [27]; Maggio et al [28]; Melchart et al [31]; Nordfeldt et al [33]; Nordqvist et al [15]
Decision-making	10 (45)	Bartonova [20]; Bowler et al [22]; Cai et al [18]; Christensen et al [23]; Das et al [24]; Kirkpatrick et al [25]; Kohler [26]; Kuijpers et al [27]; Maggio et al [28]; Melchart et al [31]

Discussion

Principal Findings

Overview

This review demonstrates that web portals are presently being used for a wide range of functions and in a variety of clinical and research settings. Ling Cai et al [18], for example, described the Lung Cancer Explorer, a database developed by researchers at the Southwestern Medical Center, University of Texas. The resource houses genomic expression and clinical data on lung cancer and is accessible to the public. Although information included on the Lung Cancer Explorer portal is a highly sophisticated technical data, it is openly accessible and can be used by a public audience attempting to learn about the genomic dimensions of lung cancer. As the third-party processing of raw consumer genetic data becomes more common [37], there may be an increase in the use of portals in the form of Lung Cancer Explorer by members of the public. Research conducted by Leanne Bowler et al [22] investigated an entirely different kind of portal, though one that combines clinical information, health research, and engages members of the public. This work details 6 web portals that provide health information to teenagers, finding that much of the information offered is of dubious quality. Two of the studies described portals for sharing genomic data, two described portals for cancer, and two for diabetes. The remaining themes included mental health, teenage health, and the intersection of health and the environment. We found that therapeutic portalsweb tools that facilitate the sharing of research findings with clinicians, researchers, and patientsare an emerging trend in health care, although they have not yet received widespread attention in the literature.

Notably, we have also found that therapeutic portals have generally not received attention as a phenomenon that is conceptually distinct from the emergence of the patient portal. An extensive literature has been published on the development and use of patient portals in recent years [38], as well as on their capacity to promote patient empowerment and self-care [39]. A growing consensus suggests that the use of a patient portal tends to positively influence health outcomes for patients with chronic disease, despite ongoing concerns about the willingness of such patients to engage with web portals [40]. Patient portals are generally thought to reflect a positive development in health care [41], even as they are accompanied by a range of pressing challenges. The popularization of patient portals has been associated, for example, with an increasing concern about the protection of sensitive or confidential patient data [42]. Our review indicates that web portals are being used for functions that generally fall outside of the scope of patient portals, as they are traditionally conceived. In particular, the increased sharing of basic research results and the expansion of portal audiences to include the general public may raise unique and unforeseen challenges. We believe that these changes may warrant a conceptual distinction between patient and therapeutic portals. As the latter becomes more prominent, it is important that they be subjected to increased scrutiny on their own terms.

Our findings suggest that therapeutic portals are likely to raise a number of important legal, ethical, and social issues. In particular, our review highlights 5 particularly prominent issues: (1) privacy and confidentiality, (2) patient health literacy, (3) equity, (4) training, and (5) decision-making. In this section, we outline how each of these issues is contemplated in our review and suggest policy and regulatory approaches that may be considered to address them.

Privacy and Confidentiality

The treatment of privacy and confidentiality issues in the sources identified in this review focused largely on the necessity of ensuring that personal data stored on a web portal is protected, and that data made available to third parties through a web resource cannot be used to identify individual research participants or patients [25]. One source discussed the importance of accounting for privacy regulations in the design of web portals that connect to personal health records [19]. Although therapeutic portals are unlikely to connect directly to personal health records, it is conceivable that they may display individualized or improperly anonymized data. A particular challenge may be determining whether a therapeutic portal is subject to local privacy regulations or is primarily governed by an internal privacy policy. Portals developed in jurisdictions with less privacy protection and made available in others may raise additional privacy considerations. Das et al [24] pointed out that certain web portals may require secure log-in procedures, including specific credentials and passwords, to protect patient information. Not surprisingly, some sources communicated empirical findings that a number of web portal users are concerned about data security and privacy [27]. Issues of privacy and confidentiality are likely to be important considerations wherever sensitive health data are processed or shared [17], including the case of therapeutic portals. With this

in view, Melchart et al [31] stress the important role project governance and strong privacy protection principles can play in advancing health research. The overall treatment of privacy and confidentiality issues in the reviewed sources helps to underscore the consensus in the literature that personal information must be carefully safeguarded, especially when such information interacts with internet-enabled tools.

Patient Health Literacy

Several articles assessed in this review described the potential health literacy and patient health literacy functions of publicly available web portals for managing and interacting with health data. Indeed, 2 of the 9 papers we found to have discussed patient health literacy, those by McKemmish et al [30] and Kuijpers et al [27], engaged explicitly with empowerment as the orienting theme of their respective contributions. McKemmish et al [30] described a web portal for sharing resources on breast cancer, writing that the *ability to access timely, relevant, and reliable information is a vital component in patient empowerment*. Access to high-quality health information can improve patient outcomes by *facilitating decision-making* and *better treatment compliance* [30]. Similarly, Kuijpers et al [27] outlined the high informational needs of cancer patients and survivors, writing that "it seems imperative to provide cancer survivors with the knowledge, skills, and motivation to positively influence their health status, which is commonly referred to as patient empowerment." Importantly, Kuijpers et al [27] noted that informational web programs can maximize their empowering potential if accompanied with face-to-face support, suggesting that web portals should not be expected to promote empowerment on their own. Other authors have identified the empowering potential of web portals in passing [15], whereas others focus on the related concept of health literacy. For example, Melholt et al [32] focused on the capacity of web portals to increase health literacy [32]. On the terms of this paper, *eHealth literacy* is conceived as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem" [32]. In this way, the relationship between health literacy and patient empowerment can easily be discerned. Tools that improve a user's understanding of their health, for example, can be expected to improve outcomes [31,32]. Sources contemplating health literacy and empowerment in this review coalesce with the view that web portals, properly implemented, may positively contribute to both of these outcomes. Critically, however, the authors note that web portals must exist within a supportive health infrastructure, offering reliable and accessible information to achieve these empowering objectives.

Equity

One possible effect of increased health literacy facilitated by web resources is the generation of greater equity in access to care. Several studies alluded to this point, although we did not find extensive discussion of health equity promotion in the sample of resources we reviewed. Melchart et al [31], for example, mentioned the importance of designing web systems that are attentive to equality of access concerns, noting that users of a service should be able to depend on receiving high-quality and equitable treatment. Similarly, Bostrom et al [21] described the development of a portal for enhancing greenspace and improving public health specifically to ensure more equitable access to recreational opportunities for underserved populations. Although this study describes a web portal significantly different from those with which we are primarily interested here, it reveals the important role that health information, collected and processed with the help of web systems, can play in advancing health equality. However, the promotion of health equity using web resources is not straightforward. Marrie et al [29] pointed out that web resource use is likely to vary substantially across sociodemographic cohorts. Technology adopters tend to be younger, more highly educated, wealthier, and have higher comorbidities. This generates a serious problem if resource developers intend for web portals to contribute to the advancements in health equity. Ensuring that such portals are accessible to everyone is almost certain to be a perennial challenge, one that is, however, necessary for achieving overarching objectives in this space, including patient empowerment and broader access to leading health research.

Training

Further challenging the implementation of novel web health resources is the issue of training. Several of the sources we reviewed detailed the particular training challenges associated with the development of tools intended to be used by clinicians, researchers, patients, and the public. For example, Das et al [24] emphasized the need to ensure that health system personnel feel adequately prepared to effectively use any new web system. Maggio et al [28] made a related point, arguing that little is known about how clinicians access and apply research findings in their practice, which has important *implications for educators designing physician training and policymakers considering public access mandates for research*. Moreover, empirical work suggests that training may be a significant barrier for clinicians to more broadly pursue the use of novel web resources in their practice. Nordfeldt et al [33] wrote that "lack of access, lack of time, and lack of opportunities for training [are] examples of causal factors preventing practitioners from adopting new technologies." Similarly, researchers accessing web systems may also require contextualizing resources that contribute to the effective use of novel tools. Bartonova et al [20], for example, stressed the importance of training regimes aimed at facilitating the work of research scientists, especially when they are encountering tools that promote interdisciplinary collaboration. In addition, web resources that bring together multiple kinds of stakeholders might also require approaches to training that include patients or the public. Kuijpers et al [27] noted that the use of web health management resources by patients or the public ought to be accompanied by sufficient training and guidance to ensure that such systems are used safely and as intended by their developers. According to Christensen et al [23], the development of web portals for certain serious medical conditions, such as in the mental health context, access to web resources might sometimes need to be accompanied by the ability to consult with trained medical professionals who are able to offer follow-up support and care [23]. Our findings in this review on training considerations are relatively

straightforward. Nevertheless, they may have important policy implications. The literature exhibits a strong consensus on the view that training systems should be commensurate with the establishment of novel web portals for health. Such training will ensure that clinicians, patients, and researchers are able to use web portals as effectively as possible, thereby maximizing potential health benefits. As portals continue to be pursued as a means to communicate research findings and general health information among stakeholders, it is critical to develop adequate strategies for training intended users.

Decision-making

Our review found that of the 22 sources, 10 (45%) discussed issues related to decision-making. Although this is somewhat an imprecise consideration, in need of greater specificity, it has potentially far-reaching legal, ethical, and policy implications for the use of web portals for health. We discovered that a number of studies gave varying degrees of attention to the ways novel web systems affect clinical and other health-related decision-making. Web resources may affect patient and public decision-making about health simply by influencing the information to which such persons have ready access. Kohler [26] made this point, stating that "information accessed through the internet enables people to make informed decisions about their life and health." Such resources may also influence medical professionals' decisions during the course of treatment [27]. The fact that web resources may affect the way health decisions are made is not, by itself, revelatory. It is, indeed, the intended function of such systems to modulate decision-making, hopefully for the more effective and efficient allocation of care resources and for improvements in individual health. However, the precise manner in which such decision-making effects are realized may be critically important. Maggio et al [28] discussed the possibility, for example, that web portals may be used to enable *shared decision-making*, in which therapeutic decisions are reached through dialog between the clinicians and their (informed) patients. Christensen et al [23] noted that certain kinds of persons may be more likely to make medical decisions based on information found on the web than others. They wrote that "consumers living with a disability or chronic disease are more frequent internet users, and more likely to state that they made health decisions based on information found on the web." Bowler et al [22] similarly found that decisions made by teenage health portal users are often *based on socially constructed content rather content provided by experts in the field*. Although portals have important implications for the manner in which medical decisions are made, they might raise questions about who makes decisions in complex health contexts. According to Das et al [24], portals accessed by clinicians should consider whether systems are established to ensure adequate decisional support and interdisciplinary interaction. Where web portals are used in the course of clinical decision-making, it is important to ensure that the provided information is supported by appropriate expertise in the circumstance. Health portal developers may also be attentive to maintaining a balance between the aim of empowering individual patients to make decisions about their own health while protecting their capacity to access high-quality clinical support and expertise. Our findings suggest that web portals may make medical decision-making significantly more efficient for numerous stakeholders, but it is critical to ensure that such decisional resources are well calibrated to the needs and capacities of the individuals using them.

Strengths and Limitations

This review attempts to collect preliminary data concerning proposals to implement therapeutic portals. This research encountered 3 important limitations that should be addressed in future work. First, the search terms employed in this review may not have been maximally targeted to the phenomenon under consideration. The reason for this is that the precise area of research with which we were interested remains relatively gestational and the terminology surrounding web portals that incorporate health research and patient perspectives has not yet been settled. It may also be that a number of therapeutic portals under development have not been described in published articles. This being the case, we hope that it will become increasingly commonplace for scholars and researchers to publish their experiences and perspectives on web portal design and implementation. Second, we did not specifically seek to include gray literature sources in this review. It is highly likely that some sources on therapeutic portals have been communicated outside of traditional academic settings. As our review focused on articles published in peer-reviewed periodicals, we did not assess any potentially pertinent gray literature sources. Future research on this topic should be expanded to include consideration of published works outside of traditional academic settings. Third, this review was limited by language and jurisdiction. Each of the 22 sources we reviewed was published in English and obtained from the English search databases. Moreover, most sources were written by scholars working in North America. These factors likely limited the scope of our search. Future research should explicitly engage with research published in languages other than English and by scholars in a more diverse set of jurisdictions. Despite these limitations, we believe that this study provides a useful overview of the landscape for therapeutic portals, including their specific potential objectives and legal, ethical, and policy challenges they are likely to face.

Conclusions

This paper communicates the findings of a scoping literature review on the potential development and implementation of what we have called therapeutic portals, web systems that facilitate the sharing of health information among researchers, clinicians, patients, and the public. Such portals differ from patient portals in several important respects. First, the role of researchers directly communicating novel findings with clinicians and the public could be significantly enhanced. This approach could help promote the growth of precision medicine as a response to complex diseases. The main aim of this study was to assess the primary legal, ethical, and policy issues associated with the development and use of web portals that facilitate information sharing between researchers, clinicians, patients, and the general public. Our findings indicate that *therapeutic portals* are a potentially promising but largely underexplored area of study. Of the 22 studies reviewed, 8 (36%) contemplated therapeutic portals were defined. The

portals we reviewed took a variety of forms and covered a range of medical issues. The absence of extensive literature on portals that specifically integrate uses from clinical, research, and public settings indicate that such systems have not been implemented on a large scale. We found that developing therapeutic portals would raise a number of important legal, ethical, and social issues, including those surrounding privacy and confidentiality, patient health literacy, equity, training, and decision-making. Each of these considerations should be given careful attention through the implementation of robust policy frameworks for managing the functions of novel portals, especially those that are of a form that has not yet been widely adopted. Further research is critical to better understand the perspectives of clinicians, researchers, patients, and members of the public on the contours of possible therapeutic portals. Although web portals that interface the numerous vital constituents in health research are, in principle, highly promising, any novel system must account for the various legal, ethical, and social challenges that they will present.

It is noteworthy that, although several of the papers included in the final review are several years post publication, those that more recently entered the literature express concern for the same kinds of issues and potential concerns. This indicates that the policy considerations identified here have not yet been fully addressed in policy development over the last decade. This underscores the responsibility for developers of the next generation of web portals to rededicate themselves to the challenge of implementing web tools that account for the serious legal, ethical, and social issues identified in this study.

This review is situated within a broader discussion in the literature on patients and other portals for managing health data and sharing information with patients and the public. The rapidly expanding use of web systems has generally been held out as a success that has contributed to improved outcomes and greater patient empowerment [43,44]. Our results suggest that web portals and systems are a promising mechanism for improving patient engagement, facilitating rapid research translation, and ultimately improving outcomes. There is great promise for portals that explicitly promote the sharing of basic research findings and engage with patients and the broader public. Portals of this kind would need to engage across a varied set of stakeholders with varying needs and levels of expertise. Further policy and empirical research will be necessary to develop strategies that are responsive to the unique challenges that such systems would raise. As web portals become increasingly important mechanisms for sharing health research with clinicians, patients, and the public, it is vital that these developments are met with ethical and conceptual scrutiny. The therapeutic portals presented in this paper may become a more widespread feature of precision and translational medicine. Our findings suggest that web portals are already being used to disseminate research results among clinicians, patients, and the public. However, much of the ethical and conceptual debate is framed in terms of the patient portal, a concept that does not adequately reflect the potentially broader scope of therapeutic portals. It may be useful to clarify this distinction in future research and to underscore the unique ethical, legal, and policy challenges raised when web systems are used as a platform for disseminating research to as wide an audience as possible.

Acknowledgments

This work was supported by the Government of Canada through Genome Canada and the Ministère de l'économie et de l'innovation du Québec through Génome Québec (grant 13528). MHZ acknowledges the generous support of the Fonds de recherche du Québec–Santé, Junior 1 Research Scholar program.

Conflicts of Interest

None declared.

References

1. Doyle D. The internet and medicine: past, present, and future. Yale J Biol Med 1996;69(5):429-437. [Medline: 9381738]
2. Robinson A. Revolutionizing 21st century medicine with consumer-based diagnostics and the internet. J Am Physicians Surg 2007;12(1):14 [FREE Full text]
3. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J 2012 Aug;7(8):992-1001 [FREE Full text] [doi: 10.1002/biot.201100306] [Medline: 22815171]
4. Klein DM, Fix GM, Hogan TP, Simon SR, Nazi KM, Turvey CL. Use of the blue button online tool for sharing health information: qualitative interviews with patients and providers. J Med Internet Res 2015 Aug 18;17(8):e199 [FREE Full text] [doi: 10.2196/jmir.4595] [Medline: 26286139]
5. Davoody N, Koch S, Krakau I, Hägglund M. Accessing and sharing health information for post-discharge stroke care through a national health information exchange platform - a case study. BMC Med Inform Decis Mak 2019 May 03;19(1):95. [doi: 10.1186/s12911-019-0816-x] [Medline: 31053141]
6. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of big data in healthcare: an overview of the european initiatives. Eur J Public Health 2019 Oct 01;29(Supplement_3):23-27 [FREE Full text] [doi: 10.1093/eurpub/ckz168] [Medline: 31738444]
7. Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

8. Mangino D, Danis M. Sharing ethics consultation notes with patients through online portals. AMA J Ethics 2020 Sep 01;22(9):784-791 [FREE Full text] [doi: 10.1001/amajethics.2020.784] [Medline: 33009775]
9. Daskalakis DC. The electronic health record and patient portals in HIV medicine: pushing the boundaries of current ethics and stigma. Camb Q Healthc Ethics 2017 Apr;26(2):332-336. [doi: 10.1017/S096318011600092X] [Medline: 28361729]
10. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Mar;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
11. Lavallée VP, Krosl J, Lemieux S, Boucher G, Gendron P, Pabst C, et al. Chemo-genomic interrogation of CEBPA mutated AML reveals recurrent CSF3R mutations and subgroup sensitivity to JAK inhibitors. Blood 2016 Jun 16;127(24):3054-3061 [FREE Full text] [doi: 10.1182/blood-2016-03-705053] [Medline: 27034432]
12. Black L, Avard D, Zawati M, Knoppers B, Hébert J, Sauvageau G, Leucegene Project. Funding considerations for the disclosure of genetic incidental findings in biobank research. Clin Genet 2013 Nov;84(5):397-406 [FREE Full text] [doi: 10.1111/cge.12190] [Medline: 23662709]
13. 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]
14. Mays N, Roberts E, Popay J. Synthesizing research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Studying the Organisation and Delivery of Health Services: Research Methods. London: Routledge; 2001:188-220.
15. Nordqvist C, Hanberger L, Timpka T, Nordfeldt S. Health professionals' attitudes towards using a Web 2.0 portal for child and adolescent diabetes care: qualitative study. J Med Internet Res 2009 Apr 06;11(2):e12 [FREE Full text] [doi: 10.2196/jmir.1152] [Medline: 19403464]
16. Kaiser J. A new portal for patient data. Science 2018 Jul 20;361(6399):212. [doi: 10.1126/science.361.6399.212] [Medline: 30026205]
17. Li R, Scott J, Rockhold F, Hill N, Wood J, Sim I. Moving data sharing forward: the launch of the vivli platform. National Academy of Medicine. 2018 Nov 19. URL: https://nam.edu/moving-data-sharing-forward-the-launch-of-the-vivli-platform/ [accessed 2021-09-29]
18. Cai L, Lin S, Girard L, Zhou Y, Yang L, Ci B, et al. LCE: an open web portal to explore gene expression and clinical associations in lung cancer. Oncogene 2019 Apr;38(14):2551-2564 [FREE Full text] [doi: 10.1038/s41388-018-0588-2] [Medline: 30532070]
19. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
20. Bartonova A. How can scientists bring research to use: the HENVINET experience. Environ Health 2012 Jun 28;11 Suppl 1:S2 [FREE Full text] [doi: 10.1186/1476-069X-11-S1-S2] [Medline: 22759502]
21. Bostrom HE, Shulaker B, Rippon J, Wood R. Strategic and integrated planning for healthy, connected cities: chattanooga case study. Prev Med 2017 Feb;95 Suppl:115-119. [doi: 10.1016/j.ypmed.2016.11.002] [Medline: 27845157]
22. Bowler L, Hong W, He D. The visibility of health web portals for teens: a hyperlink analysis. Online Inform Rev 2011 Jun 21;35(3):443-470 [FREE Full text] [doi: 10.1108/14684521111151469]
23. Christensen H, Murray K, Calear AL, Bennett K, Bennett A, Griffiths KM. Beacon: a web portal to high-quality mental health websites for use by health professionals and the public. Med J Aust 2010 Jun 07;192(S11):40-44. [doi: 10.5694/j.1326-5377.2010.tb03692.x] [Medline: 20528708]
24. Das A, Faxvaag A, Svanæs D. The impact of an eHealth portal on health care professionals' interaction with patients: qualitative study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
25. Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, ClinGen Resource. GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. Hum Mutat 2015 Oct;36(10):974-978 [FREE Full text] [doi: 10.1002/humu.22838] [Medline: 26178529]
26. Kohler S. Can internet access growth help reduce the global burden of noncommunicable diseases? Online J Public Health Inform 2013;5(2):221 [FREE Full text] [doi: 10.5210/ojphi.v5i2.4667] [Medline: 23923103]
27. Kuijpers W, Groen WG, Loos R, Oldenburg HS, Wouters MW, Aaronson NK, et al. An interactive portal to empower cancer survivors: a qualitative study on user expectations. Support Care Cancer 2015 Sep;23(9):2535-2542 [FREE Full text] [doi: 10.1007/s00520-015-2605-0] [Medline: 25876157]
28. Maggio LA, Moorhead LL, Willinsky JM. Qualitative study of physicians' varied uses of biomedical research in the USA. BMJ Open 2016 Nov 21;6(11):e012846 [FREE Full text] [doi: 10.1136/bmjopen-2016-012846] [Medline: 27872121]
29. Marrie RA, Leung S, Tyry T, Cutter GR, Fox R, Salter A. Use of eHealth and mHealth technology by persons with multiple sclerosis. Mult Scler Relat Disord 2019 Jan;27:13-19. [doi: 10.1016/j.msard.2018.09.036] [Medline: 30296732]
30. McKemmish S, Manaszewicz R, Burstein F, Fisher J. Consumer empowerment through metadata-based information quality reporting: the breast cancer knowledge online portal. J Am Soc Inf Sci 2009 Sep;60(9):1792-1807. [doi: 10.1002/asi.21096]
31. Melchart D, Eustachi A, Gronwald S, Wühr E, Wifling K, Bachmeier BE. Introduction of a web portal for an individual health management and observational health data sciences. Patient Relat Outcome Meas 2018;9:183-196 [FREE Full text] [doi: 10.2147/PROM.S157410] [Medline: 29950912]

32. Melholt C, Joensson K, Spindler H, Hansen J, Andreasen JJ, Nielsen G, et al. Cardiac patients' experiences with a telerehabilitation web portal: implications for eHealth literacy. Patient Educ Couns 2018 May;101(5):854-861. [doi: 10.1016/j.pec.2017.12.017] [Medline: 29305064]
33. Nordfeldt S, Ängarne-Lindberg T, Berterö C. To use or not to use--practitioners' perceptions of an open web portal for young patients with diabetes. J Med Internet Res 2012 Nov 09;14(6):e154 [FREE Full text] [doi: 10.2196/jmir.1987] [Medline: 23137767]
34. Rocker C, Cappelletti L, Marshall C, Meunier CC, Brooks DW, Sherer T, et al. Use of an online portal to facilitate clinical trial recruitment: a preliminary analysis of Fox Trial Finder. J Parkinsons Dis 2015;5(1):55-66. [doi: 10.3233/JPD-140522] [Medline: 25624422]
35. Sutherland JJ, Stevens JL, Johnson K, Elango N, Webster YW, Mills BJ, et al. A novel open access web portal for integrating mechanistic and toxicogenomic study results. Toxicol Sci 2019 Aug 01;170(2):296-309 [FREE Full text] [doi: 10.1093/toxsci/kfz101] [Medline: 31020328]
36. Tomlinson C, Thimma M, Alexandrakis S, Castillo T, Dennis JL, Brooks A, et al. MiMiR--an integrated platform for microarray data sharing, mining and analysis. BMC Bioinformatics 2008 Sep 18;9:379 [FREE Full text] [doi: 10.1186/1471-2105-9-379] [Medline: 18801157]
37. Moscarello T, Murray B, Reuter CM, Demo E. Direct-to-consumer raw genetic data and third-party interpretation services: more burden than bargain? Genet Med 2019 Mar;21(3):539-541 [FREE Full text] [doi: 10.1038/s41436-018-0097-2] [Medline: 29997392]
38. Antonio M, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? a scoping review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):871-883 [FREE Full text] [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
39. 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]
40. Coughlin S, Prochaska J, Williams LB, Besenyi G, Heboyan V, Goggans S, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy 2017 Apr 07;10:33-40 [FREE Full text] [doi: 10.2147/RMHP.S130431] [Medline: 28435342]
41. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
42. Bush RA, Connelly CD, Fuller M, Pérez A. Implementation of the integrated electronic patient portal in the pediatric population: a systematic review. Telemed J E Health 2016 Mar;22(2):144-152 [FREE Full text] [doi: 10.1089/tmj.2015.0033] [Medline: 26258289]
43. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
44. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, Dabbs AD. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]

Abbreviations

AML: acute myeloid leukemia

Edited by R Kukafka; submitted 11.12.20; peer-reviewed by M Avdagovska, H Namazi; comments to author 26.01.21; revised version received 17.03.21; accepted 04.08.21; published 11.11.21.

Please cite as:
Lang M, Lemieux S, Hébert J, Sauvageau G, Zawati MH
Legal and Ethical Considerations for the Design and Use of Web Portals for Researchers, Clinicians, and Patients: Scoping Literature Review
J Med Internet Res 2021;23(11):e26450
URL: https://www.jmir.org/2021/11/e26450
doi:10.2196/26450
PMID:34762055

Original Paper

Fitness Tracker Information and Privacy Management: Empirical Study

Mohamed Abdelhamid[1], PhD

Department of Information Systems, California State University, Long Beach, Long Beach, CA, United States

Corresponding Author:
Mohamed Abdelhamid, PhD
Department of Information Systems
California State University, Long Beach
1250 N Bellflower Blvd
Long Beach, CA, 90840
United States
Phone: 1 5629852361
Email: mohamed.abdelhamid@csulb.edu

Abstract

Background: Fitness trackers allow users to collect, manage, track, and monitor fitness-related activities, such as distance walked, calorie intake, sleep quality, and heart rate. Fitness trackers have become increasingly popular in the past decade. One in five Americans use a device or an app to track their fitness-related activities. These devices generate massive and important data that could help physicians make better assessments of their patients' health if shared with health providers. This ultimately could lead to better health outcomes and perhaps even lower costs for patients. However, sharing personal fitness information with health care providers has drawbacks, mainly related to the risk of privacy loss and information misuse.

Objective: This study investigates the influence of granting users granular privacy control on their willingness to share fitness information.

Methods: The study used 270 valid responses collected from Mtrurkers through Amazon Mechanical Turk (MTurk). Participants were randomly assigned to one of two groups. The conceptual model was tested using structural equation modeling (SEM). The dependent variable was the intention to share fitness information. The independent variables were perceived risk, perceived benefits, and trust in the system.

Results: SEM explained about 60% of the variance in the dependent variable. Three of the four hypotheses were supported. Perceived risk and trust in the system had a significant relationship with the dependent variable, while trust in the system was not significant.

Conclusions: The findings show that people are willing to share their fitness information if they have granular privacy control. This study has practical and theoretical implications. It integrates communication privacy management (CPM) theory with the privacy calculus model.

(J Med Internet Res 2021;23(11):e23059) doi:10.2196/23059

KEYWORDS

privacy; information sharing; fitness trackers; wearable devices

Introduction

Background

Fitness trackers are wearable devices that allow people to monitor and track activities and information related to fitness, such as distance walked and calories consumed. Fitness trackers can be stand-alone devices or integrated within a smartwatch. The device is usually connected to a mobile app that allows users to manage information and use the features of the app. The first functional fitness tracker was invented in the mid-1960s [1]. However, the rise of fitness trackers and wearable devices started about a decade ago. The first Fitbit was released in 2012 [2], and the first Apple watch was revealed in 2014 [3]. Since then, fitness trackers and wearable devices have become increasingly popular. A recent Pew Research Center study reported that 21% of Americans regularly use fitness trackers or smartwatches [4]. Similarly, in 2019, Gallup reported that more than one in four Americans use an app or device to track fitness-related activities [5].

People use fitness trackers for various reasons, but ultimately, the main reason is to get fit or maintain health [6,7]. The device/app helps users stay motivated and allows them to track progress and stay informed [8]. In recent years, many people have started seeking a healthier lifestyle and adopted technologies that motivated them to keep track of their goals [9]. This trend is largely adopted by millennials; some even called the millennials "the wellness generation" [10].

Information collected by fitness trackers and maintained by their respective systems, if shared with health care providers, could provide wider benefits to users. Sharing personal fitness data with health care providers allows physicians to better understand their patients' health lifestyle, health issues, and potential health problems. This further allows physicians to provide early recommendations and make better health assessments that could help people avoid health problems. In general, sharing personal fitness information with health care institutions may benefit individuals in many ways, such as better health outcomes and reduced cost. Researchers are starting to predict a future that encourages patients to share fitness data with their providers [11]. Another possible benefit of sharing fitness data is to conduct scientific research that could improve the health outcomes for the general public.

However, the use of technology and the sharing of personal fitness data may result in negative consequences, mainly related to privacy and security [12]. For example, loss of privacy may result from a breach of security in a health care institution. More than 93% of health care institutions have been victims of a data breach in the past 5 years [13].

This study investigates the influence of implementing granular privacy control on users' intention to share their fitness information and whether it could lead to higher user engagement in sharing fitness data. In addition, the paper investigates the motivation of individuals, captured by the privacy calculus, to share fitness information. This research provides several important contributions to the field. First, it contributes to theory by integrating communication privacy management (CPM) theory with the privacy calculus model [14]. Second, the paper provides practical and theoretical insights into how to address barriers to fitness information sharing by suggesting flexible sharing mechanisms that mitigate the impact of perceived privacy risk.

Theoretical Background and Hypothesis Development

Privacy in eHealth

The diffusion of health-tracking apps and devices is growing rapidly in the United States, both in terms of the number of people that are active end users and also in terms of the functionalities and features of the apps and devices that are currently available. Fitness and health-tracking apps are expected to have many benefits for the users, such as motivation and fitness aspiration [15]. However, with digitization come the risks of privacy and security breaches [16,17]. People's behavior with regard to sharing personal health information is negatively influenced by concerns over their privacy [18,19]. Individual-centered privacy research found that individuals are concerned about the collection, handling, and possible unauthorized access of their private information [18].

Research on health information sharing summarizes that risks of privacy invasion and information violations are drawbacks of sharing [20]. This means that individuals are hesitant to share personal health information because of possible privacy risks. Although sharing is beneficial in many cases, the risks may outweigh the benefits, and thus risk can drive adoption [18]. Therefore, granular privacy control mechanisms for information sharing may motivate individuals to share more of their fitness information. Cavusoglu et al [21] found that in the social media context, granular control motivates users to share more information because they control with whom they share information. This research covers the gap by testing for the impact of a more granular privacy control in sharing fitness tracker information.

Information Sharing

One body of literature focused on the sharing of health-related information via online and electronic sources. Prior research suggests that privacy concerns are the central obstacle to sharing of information [18,22,23]. Simon et al [24] identified privacy and security issues and lack of benefits as the main barriers to sharing of personal information.

Angst and Agarwal [18] confirmed that privacy concerns reduce the likelihood of sharing health information. Weitzman et al [25] found that patients with sensitive information are less likely to share their health information with health care providers. Likewise, Zulman et al [26] confirmed that willingness to share health information is influenced by the type of information. Bansal and Gefen [23] reported that the sensitivity of information influences individuals' decisions to share that online.

The key factors in sharing health information include the benefits of obtaining feedback related to potential health issues. According to Dimitropoulos et al [27], most people realize the benefits of sharing health information. However, they need to adapt to and manage the way their information is shared. Although Ancker et al [28] found that most respondents believe that sharing health information improves the quality of care.

Privacy Calculus Model

Information sharing and privacy have always been supplementary [29,30]. Therefore, the theoretical model of this research is guided by the privacy calculus model [14]. Studies in various contexts have used the privacy calculus model [31]. For example, Kim et al [32] used the privacy calculus model to investigate people's willingness to provide personal information in the context of the internet of things (IoT). They found that perceived benefits are a strong motivator in sharing private information. Likewise, Fox [33] investigated the influence of privacy calculus variables on individuals' intention to adopt mobile health technologies. The paper finds a stronger influence of benefits compared to risks and concerns. Abdelhamid et al [34], when examining factors associated with the sharing of health information, found that privacy is the biggest barrier to sharing.

The privacy calculus model is a good fit for this study because it deals with information sharing in scenarios where risk and benefits of sharing are involved. However, the privacy calculus model does not incorporate granular control of information into the model. Thus, this research integrates CPM, which incorporates control of information, with privacy calculus, which deals with risks and benefits. Therefore, the overall model covers the three key factors of this research: granular control, risks, and benefits.

This research adopts the theory by:

1. Applying the theory in the context of fitness trackers and wearable devices
2. Integrating CPM theory with the privacy calculus model

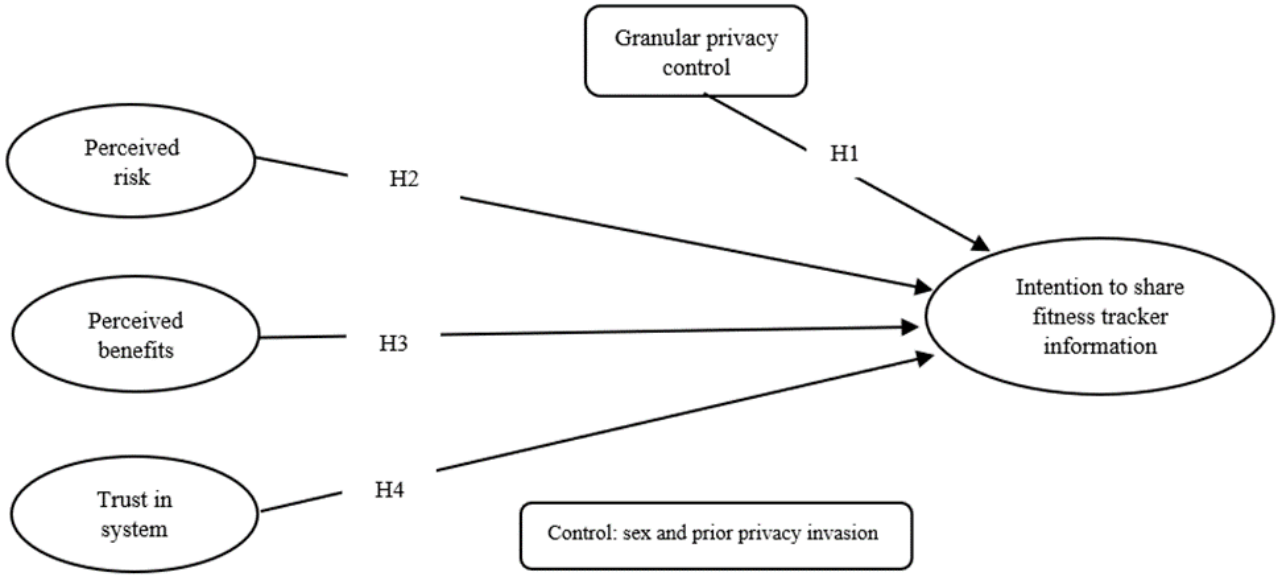
CPM

CPM theory describes the rationale behind an individual's choice to disclose or withhold private information [35].

The principles state that people believe they are the ultimate owners of their private information and that people have the right to control the course of their private information. Thus, people believe they have the right to choose with whom to share information. CPM also argues that people believe that stakeholders accessing an individual's shared information will follow current and future privacy rules. In addition, CPM debates that violated privacy rules will result in negative consequences, including mistrust and uncertainty. Petronio [35] underlined that CPM was developed to help users make necessary alterations to systems when faced with the issue of privacy.

According to the breach data reported to the Department of Health Human Services, the number of unauthorized access/disclosure breaches tripled in 2017 versus 2012 [36]. Dhopeshwarkar et al [37] noted that most people want to know who is viewing their information. Angst et al [38] suggested that institutional factors and IT investments affect the likelihood of breaches in health care organizations. This research investigates the impact of a more granular privacy control on the intention to share data related to fitness activities with health care providers. Figure 1 shows the conceptual model of this study.

Figure 1. Conceptual model. H: Hypothesis.



Privacy Control

When faced with a choice, individuals typically choose the option with the premier value after assessing likely risks [39]. In the context of fitness and health and from an individual perspective, there may be a desire to protect specific fitness information from certain providers or users. Caine and Hanania [40] suggested that people prefer to share particular health-related information with specific recipients.

The definition of privacy refers to the right that people have in choosing the information they want to share and with whom to share it [41]. Information control and information disclosure have a positive relationship in various contexts [42,43]. For instance, Cavusoglu et al [21] examined the causal effect of granting Facebook users more control on information-sharing behavior. They found that privacy control increases the open release of information. In other words, when the sharing decision is universal, people might not share some information with anyone. That is mainly because some people want to prevent others from seeing that information. As a result, the decision led to withholding the information from everyone. However, when individuals are given more control over their information, they will share some information with some people.

In addition, Slovic [44] found that on average, individuals are ready to take more risks when they are in control. The paper suggested that improved control increases an individual's willingness to participate in that behavior. Likewise, Brandimarte et al [45] reported that individuals are more willing to reveal sensitive information when they have more control over what is being shared.

The first hypothesis of this study is as follows:

Hypothesis 1: Granular privacy control will yield a higher intention to share fitness information with health care providers.

Perceived Risk

The concept of perceived risk has been studied in many contexts in which individuals may face a risky decision. A considerable number of studies have established the link between perceived risk and information disclosure, in general [46-48]. Dinev and Hart [14] defined perceived risk as the perceived risk of unprincipled behavior related to the sharing of personal information. In the context of fitness tracker information, the risk includes selling the information to a third party, misuse, and unauthorized sharing. In 2015, the National Telecommunications and Information Assurance (NTIA) surveyed approximately 40,000 participants [49]. The report stated that more than 50% of users had limited their online activities due to concerns about the privacy of their information.

The second hypothesis of this study is as follows:

Hypothesis 2: Perceived risk will have a negative influence on the intention to share fitness information with health care providers.

Perceived Benefits

When people face a decision that involves sharing of private fitness information, they usually assess the risks and benefits of sharing to make an informed decision. Many benefits result from sharing fitness information with health care providers. Some of those benefits directly influence the individuals sharing the fitness information. For example, fitness information can help doctors make better health assessments with regard to the person sharing the information. This allows doctors to make better recommendations. The individual may then benefit from better health outcomes, in general.

In the context of fitness, perceived benefit is defined as the perceived value that individuals attach to sharing personal fitness information with health care providers. In general, perceived benefits have been associated with information sharing in various contexts, including health care [50]. For example, Wang et al [51] reported that patients find that improved health care quality and convenience are among the benefits of sharing personal health information. Likewise, Zhang et al [52] found that sharing health information in online communities is associated with benefits for users. They argue that in online communities, the benefits are informational and emotional support. Morris et al [53] proposed a design of a mobile information-sharing system for emergency rooms. They found that sharing can be beneficial for physicians in terms of reducing information-seeking time and stress. This could result in better care for patients. As a result of sharing, patients may be able to avoid a serious problem.

The third hypothesis of this study is as follows:

Hypothesis 3: Perceived benefits will have a positive influence on the intention to share fitness information with health care providers.

Trust in a System

Trust in a system is defined as the extent to which individuals are confident that systems will handle their information securely and reliably [14]. The prior literature has established a positive relationship between trust in a system and engagement with the system [54,55]. The perception of trust can be linked with the system itself or with the system's capability to protect information from people who breach the system to misuse information. The perspective of information misuse escalates when private fitness information is exchanged from one system to another. Gefen et al [56] state that the relationship between trust in a system and the intention to use that system becomes more significant when engagement includes the possibility of risk consequences.

The fourth hypothesis of this study is as follows:

Hypothesis 4: Trust in the system will have a positive influence on the intention to share fitness information with health care providers.

Methods

Data Collection

This study uses scenario-based survey data collection through Amazon Mechanical Turk (MTurk). A survey-based approach has been used in many studies in the context of health care IT to understand individuals' perceptions related to information sharing [18,20]. This study aims to understand individuals' perception and intentions as they relate to sharing fitness tracker information. Thus, a survey-based study is adequate. Many studies in the health care field have used Amazon MTurk to collect data [57]. Online data collection is relevant to this study for many reasons. First, fitness trackers are used by the general public and not restricted to a certain occupation or demographic. Second, most people in the United States have regular access to the internet [58]. Third, online data collection, compared with convenience sampling, allows for reaching participants outside the researcher's geographic area.

Participants were asked to participate in a study related to fitness tracker information. After reading the consent form, the participants were asked to start the survey. The first question asks participants whether they have owned a fitness tracker. The survey ended for those who indicated they did not own a fitness tracker. The participants were paid USD 0.50 for completing the survey. The participants were randomly assigned to one of two groups (granular or universal). Under the granular privacy scenario, the participants were exposed to a scenario where they could select what fitness information to share and with whom. Next, they were asked how likely they are to share the information under such a scenario. In universal sharing, the participants were told that they could not exclude specific information from sharing. Next, they were asked to indicate their sharing intentions under this scenario. All participants answered the same questions related to independent and control variables. Control variables were sex and prior privacy invasion, included to follow the design of Angst et al [18], within the health care context.

Data Summary

The final data analysis included 270 valid and complete responses. The participants had to answer questions related to the independent variables, as shown in Figure 1. Next, each participant was exposed to one of two sharing settings. Finally, the participants indicated their intention to share their fitness information with health care providers, depending on the sharing setting to which they were exposed. Qualtrics settings allowed for random assignment, while keeping the number of participants in the two groups similar. In total, 137 participants were assigned to the granular privacy sharing option (select what fitness information to disclose and share and whom to share it with), and 133 were assigned to the universal sharing option (share all personal fitness information with all providers). Of the 270 participants, 77.8% (n=211) were male, and the rest (22.2%, n=59) were female. The majority of the participants were between the age of 25 and 34 years (163 participants), the second-largest group was 35-44 years old (39 participants), 33 participants were between the age of 18 and 24 years, 21 participants were between the age of 45 and 54 years, and 14 participants were 55 years old or older.

Measurement Model Assessment

SAS software version 9.4 was used to decode the data, and IBM AMOS version 25 was used to run the analysis. Confirmatory factor analyses were used to evaluate the measurement model (Table 1) using all 270 participants. All variables in this study were adapted from prior research (see Table A1 of Multimedia Appendix 1 for measurement items). All latent variables were measured on a 5-point Likert scale. The results of the measurement model showed a good fit [59]. All factor loadings for the latent variables were relatively strong and significant. The comparative fit index (CFI)=.964, root-mean-square error of approximation (RMSEA)=.059, Tucker-Lewis index (TLI)=.953, and χ^2/df=1.724. These results provided evidence of the validity of the constructs.

Table 1. Measurement model.

Latent variable	Item	Loadings	Corrected item–total correlation	Construct reliability	Variable inflation factor
Intention to share (dependent variable)	INT_1	0.82	0.730	0.863	NA[a]
	INT_2	0.782	0.702		
	INT_3	0.867	0.778		
Perceived risk	PR_1	0.804	0.619	0.764	1.047
	PR_2	0.633	0.550		
	PR_3	0.718	0.612		
Perceived benefits	PB_1	0.7	0.616	0.758	1.003
	PB_2	0.69	0.577		
	PB_3	0.753	0.576		
Trust in the system	TR_1	0.733	0.545	0.721	1.029
	TR_2	0.591	0.511		
	TR_3	0.712	0.569		
Prior experience with privacy invasion (control)	PI_1	0.847	0.747	0.850	1.015
	PI_2	0.785	0.699		
	PI_3	0.794	0.710		

[a]NA: not available.

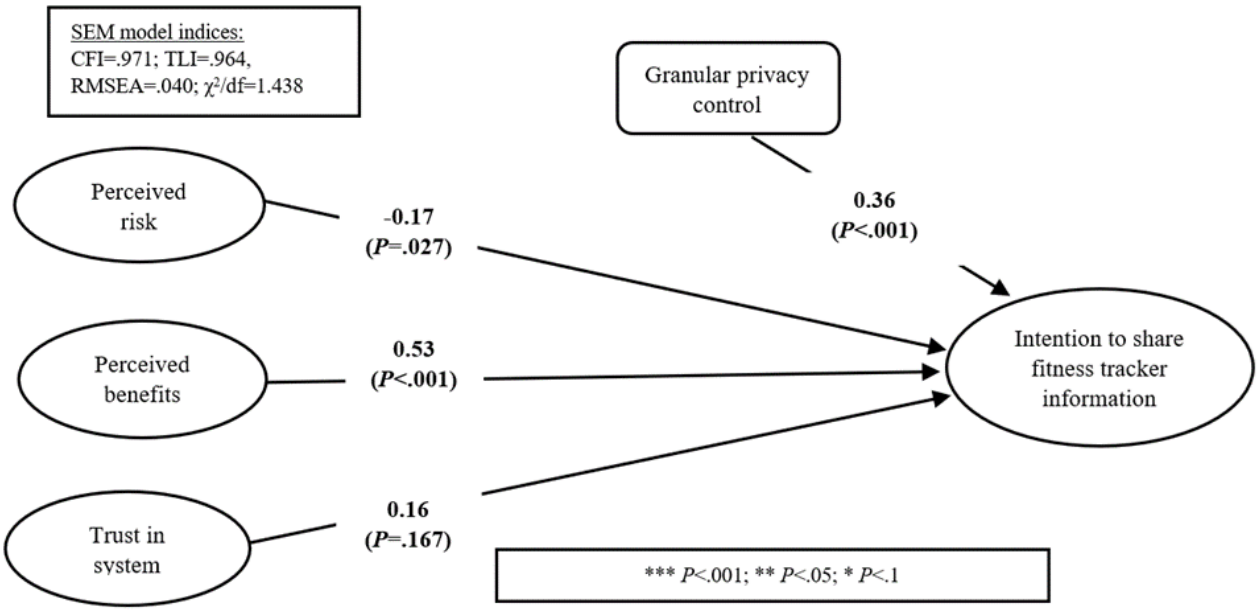
The reliability of constructs was assessed by calculating the composite reliability (CR). The reliability scores for all constructs in the conceptual model exceeded the threshold of 0.7, which indicates strong reliability. The CR scores ranged from 0.721 to 0.863 (see Table 1). In addition, the corrected item–total correlation for each item was calculated based on the construct to which it belonged. All values exceeded the minimum cutoff of 0.5 [60]. Furthermore, the variance inflation factor (VIF) was calculated for each of the independent variables in the measurement model to check for multicollinearity. All VIF values were way below the threshold score of 10. Therefore, there was no evidence for the existence of multicollinearity between variables in this study.

Results

SEM Results

Figure 2 shows the results of structural equation modeling (SEM). The model explained 60.4% of the variance (R^2) in the intention to share.

Figure 2. SEM results. CFI: comparative fit index; RMSEA: root-mean-square error of approximation; SEM: structural equation modeling; TLI: Tucker-Lewis index.



Granular Privacy Control

Hypothesis 1 states that increased control results in a higher intention to share. Findings supported this result. The path coefficient for granular privacy control was positive and significant (β_{GPC}=.36, *P*<.001), indicating that granular privacy control yields a higher willingness to share personal fitness-related information with health care providers.

Perceived Risk

Hypothesis 2 proposes a positive relationship between negative perceived risk and the intention to share fitness information. The results provided evidence to support this hypothesis (β_{PR}=–.17, *P*=.027), which confirms the impact of the possible risk that involves sharing information via systems and with others.

Perceived Benefits

Hypothesis 3 argues that perceived benefits will yield a higher intention to share fitness information with health care providers. The estimate for this relationship was positive and significant (β_{PB}=.53, *P*<.001), which provides support for the hypothesis. The magnitude of the influence was the highest among all variables, which confirms the importance of benefits for individuals to be willing to share their fitness information.

Trust in a System

Finally, hypothesis 4 states that trust in the system will have a positive influence on the intention to share personal fitness information with health care providers. However, the results did not provide support for this hypothesis. The estimate was positive but not significant. Thus, trust does not seem to be an issue or a barrier to information sharing in this context. Prior invasion of privacy and sex were used as control variables. It is expected that individuals' perceptions are influenced by the prior invasion of privacy. Both variables were positively significant.

Discussion

Principal Findings

Privacy calculus theory highlights that people weigh benefits and risks when making decisions related to sharing or disclosing personal information. This research finds that both risk and benefits have a significant influence on the intention to share fitness data (see Figure 2). These findings are similar to the findings of other research in the health care and cybersecurity context [18,51]. However, the benefits of sharing are more influential compared to the risks, at least in this sample. This result is positive for health care providers, researchers, and those who need the information to improve population health. User benefits include improved health care quality, more accurate information, more convenience, and better communication. In addition, people who share their fitness information with health care providers may be able to avoid serious problems by allowing the providers to detect problems early. This could result in avoiding increases in health insurance premiums for individuals who share their information.

Cavusoglu et al [21] showed that granular privacy control motivates Facebook users to share more content because they are able to control the content they can share and with whom they can share it. This is in line with the results of this research that show that granular privacy control could motivate people to share their fitness tracker information. Prior research [18,20] has shown that sharing health-related information is perceived by individuals to be risky. The results of this study confirm that sharing fitness information is also perceived to be risky. However, the benefits seem to outweigh the risks.

In addition, this paper integrated CPM with the privacy calculus model. CPM defines the motivation behind individuals' choice

to share or withhold private information. In this paper, participants were randomly assigned to scenarios (granular and universal sharing). Findings showed that granting people greater privacy control acts as a persuasive mechanism to motivate more people to participate in sharing their fitness information. Thus, individuals can engage in behaviors that may improve their well-being, while taking actions to protect their private data. This is an implication for policymakers to enforce granular privacy sharing settings that will allow individuals to participate in such systems and, in turn, observe better health outcomes. In addition, a higher participation rate will allow those applications to be sustainable as they enable more people to benefit from the system.

Trust in the system does not have a significant relationship with fitness information sharing. This finding requires further investigation because it goes against the hypothesis and previous research as it relates to information sharing. However, several explanations are plausible. For example, in this era, most people use apps and systems all the time. Thus, the concept of general trust in the system starts to vanish as systems become part of our daily work and personal routines.

Limitations and Future Work

This study had several limitations. First, the dependent variable was the intention to share fitness information with health care providers and not actual behavior. However, previous studies have indicated that intention is a strong predictor of actual behavior [61]. Another limitation was that the data were collected online through Amazon MTurk. This could also be associated with selection bias. However, many studies in the health care field have used online data collection methods. In addition, after the COVID-19 situation, online data collection is expected to become more prominent. Furthermore, the integration of fitness apps and systems of health care providers has not been adopted yet, at least not on a large scale. Future work will focus on other aspects of application design and privacy and security settings.

Conclusion and Contribution

The findings of this study have many implications for practice and the literature. Individuals, generally, choose to share specific information with specific health care providers. Viewed from a privacy perspective, enforcement of granular privacy settings lessens the perceived risk by giving individuals a greater sense of assurance regarding their personal fitness information. This research finds that on average, people are likely to share their fitness information when applications empower them with more control. That is because people naturally prefer to avoid risk. Granular privacy control offers people the ability to mitigate risk. This step will increase their willingness to participate in sharing personal fitness information.

This paper illustrates how providing individuals with granular privacy control can lead to improvement in sharing of fitness information. This could result in improved health outcomes for individuals and the general public. Granular privacy control allows individuals to mitigate the perceived risk involved in the universal sharing of all fitness information.

In general, the perceived risk remains a major barrier to information sharing, even with regard to fitness information. The introduction of granular privacy control could mitigate the negative impact of perceived risk. On the positive side, perceived benefits show the strongest influence on the intention to share fitness information. This indicates that individuals attach sharing fitness information to many benefits. The magnitude of the perceived benefit coefficient is three times stronger than the coefficient of perceived risk. This also has implications for the need for such integration between fitness apps and health care systems. Policymakers may want to consider establishing policies and rules that govern the sharing process.

This research contributes to theory by integrating the privacy model and CPM theory in the context of fitness information sharing. In addition, the study adds to theory by highlighting the impact of granular privacy control on the intention to share fitness information.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement items.
[DOCX File , 17 KB - jmir_v23i11e23059_app1.docx]

References

1. Douglas-Walton J. A Study of Fitness Trackers and Wearables. URL: https://tinyurl.com/3efv4mte [accessed 2021-07-15]
2. Comstock J. Eight Years of Fitbit News Leading up to Its Planned IPO. 2015. URL: https://tinyurl.com/fhkdrxxa [accessed 2020-07-30]
3. Dormehl L. Today in Apple History: It's Time for Apple Watch. URL: https://tinyurl.com/nsdc4fem [accessed 2021-04-25]
4. Vogels E. About One-in-Five Americans Use a Smart Watch or Fitness Tracker. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2020-06-30]
5. Mccarthy J. One in Five U.S. Adults Use Health Apps, Wearable Trackers. 2019. URL: https://news.gallup.com/poll/269096/one-five-adults-health-apps-wearable-trackers.aspx [accessed 2020-06-30]

6. Asimakopoulos S, Asimakopoulos G, Spillers F. Motivation and user engagement in fitness tracking: heuristics for mobile healthcare wearables. Informatics 2017 Jan 22;4(1):5. [doi: 10.3390/informatics4010005]
7. Wu Q, Sum K, Nathan-Roberts D. How fitness trackers facilitate health behavior change. 2016 Sep 15 Presented at: The Human Factors and Ergonomics Society Annual Meeting; September 19-23, 2016; WA p. 1068-1072. [doi: 10.1177/1541931213601247]
8. Naglis M, Bhatiasevi V. Why do people use fitness tracking devices in Thailand? An integrated model approach. Technol Soc 2019 Aug;58:101146. [doi: 10.1016/j.techsoc.2019.101146]
9. Wellness Is The New Luxury: Is Healthy And Happy The Future Of Retail?. 2017. URL: https://www.forbes.com/sites/deborahweinswig/2017/06/30/wellness-is-the-new-luxury-is-healthy-and-happy-the-future-of-retail/#14f6f0338323 [accessed 2020-06-15]
10. Nermoe K. Millennials: The "Wellness Generation". 2018. URL: https://news.sanfordhealth.org/sanford-health-plan/millennials-wellness-generation/ [accessed 2020-06-15]
11. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
12. Gabriele S. Understanding fitness tracker users' security and privacy knowledge, attitudes and behaviours. 2020 Apr 30 Presented at: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376651]
13. Osborne C. The Latest Healthcare Data Breaches in 2019/2020. URL: https://portswigger.net/daily-swig/the-latest-healthcare-data-breaches [accessed 2021-07-30]
14. Dinev T, Hart P. An extended privacy calculus model for e-commerce transactions. Inf Syst Res 2006 Mar;17(1):61-80. [doi: 10.1287/isre.1060.0080]
15. Birelin A. The Benefits of Logging Workouts into a Fitness App. 2017. URL: https://www.nifs.org/blog/the-benefits-of-logging-workouts-into-a-fitness-app [accessed 2020-06-15]
16. Perez A. Use a Fitness App to Track Your Workouts? Your Data May Not Be as Protected as You Think. 2019. URL: https://www.usatoday.com/story/sports/2019/08/16/what-info-do-fitness-apps-keep-share/1940916001/ [accessed 2020-04-15]
17. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
18. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Quarterly 2009;33(2):339. [doi: 10.2307/20650295]
19. Abdelhamid M, Kisekka V, Samonas S. Mitigating e-services avoidance: the role of government cybersecurity preparedness. ICS 2019 Mar 11;27(1):26-46. [doi: 10.1108/ics-02-2018-0024]
20. 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]
21. Cavusoglu H, Phan TQ, Cavusoglu H, Airoldi EM. Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. Inf Syst Res 2016 Dec;27(4):848-879. [doi: 10.1287/isre.2016.0672]
22. Li H, Gupta A, Zhang J, Sarathy R. Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract. Decis Support Syst 2014 Jan;57:376-386. [doi: 10.1016/j.dss.2012.10.043]
23. Bansal G, Zahedi, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
24. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30-e41 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
25. Orehek J, Charpin D, Gayrard P, Grimaud C. Bronchodilator properties of a vasodilator: cetiedil. Nouv Presse Med 1976 Jun 19;5(25):1577-1578. [Medline: 6947]
26. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810 [FREE Full text] [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
27. Dimitropoulos L, Patel V, Scheffler SA, Posnack S. Public attitudes toward health information exchange: perceived benefits and concerns. Am J Manag Care 2011 Dec;17(12 Spec No):SP111-SP116 [FREE Full text] [Medline: 22216769]
28. Ancker JS, Edwards AM, Miller MC, Kaushal R. Consumer perceptions of electronic health information exchange. Am J Prev Med 2012 Jul;43(1):76-80. [doi: 10.1016/j.amepre.2012.02.027] [Medline: 22704751]
29. Gerlach J, Widjaja T, Buxmann P. Handle with care: How online social network providers' privacy policies impact users' information sharing behavior. J Strateg Inf Syst 2015 Mar;24(1):33-43. [doi: 10.1016/j.jsis.2014.09.001]
30. Smith HJ, Dinev T, Xu H. Information privacy research: an interdisciplinary review. MIS Quarterly 2011;35(4):989. [doi: 10.2307/41409970]

31. Sun Y, Wang N, Shen X, Zhang JX. Location information disclosure in location-based social network services: privacy calculus, benefit structure, and gender differences. Comput Hum Behav 2015 Nov;52:278-292. [doi: 10.1016/j.chb.2015.06.006]
32. Kim D, Park K, Park Y, Ahn J. Willingness to provide personal information: perspective of privacy calculus in IoT services. Comput Hum Behav 2019 Mar;92(4):273-281. [doi: 10.1016/j.chb.2018.11.022]
33. Fox G. "To protect my health or to protect my health privacy?" A mixed‐methods investigation of the privacy paradox. J Assoc Inf Sci Technol 2020 May 20;71(9):1015-1029. [doi: 10.1002/asi.24369]
34. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
35. Petronio S. Translational research endeavors and the practices of communication privacy management. J Appl Commun Res 2007 Aug;35(3):218-222. [doi: 10.1080/00909880701422443]
36. McLeod A, Dolezel D. Cyber-analytics: modeling factors associated with healthcare data breaches. Decis Support Syst 2018 Apr;108:57-68. [doi: 10.1016/j.dss.2018.02.007]
37. Dhopeshwarkar RV, Kern LM, O'Donnell HC, Edwards AM, Kaushal R. Health care consumers' preferences around health information exchange. Ann Fam Med 2012 Sep 10;10(5):428-434 [FREE Full text] [doi: 10.1370/afm.1396] [Medline: 22966106]
38. Angst CM, Block ES, D'Arcy J, Kelley K. When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. MISQ 2017 Mar 3;41(3):893-916. [doi: 10.25300/misq/2017/41.3.10]
39. Tversky A, Fox CR. Weighing risk and uncertainty. Psychol Rev 1995;102(2):269-283. [doi: 10.1037/0033-295x.102.2.269]
40. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 01;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]
41. Westin A. Privacy and freedom. In: Washington and Lee Law Review. NY: Atheneum; 1967:166.
42. Laugesen J, Hassanein K. Adoption of personal health records by chronic disease patients: a research model and an empirical study. Comput Hum Behav 2017 Jan;66:256-272. [doi: 10.1016/j.chb.2016.09.054]
43. Abdelhamid M. Greater patient health information control to improve the sustainability of health information exchanges. J Biomed Inform 2018 Jul;83:150-158 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.002] [Medline: 29894811]
44. Slovic P. The Perception of Risk: Risk, Society and Policy. London: Earthscan; 2000:473.
45. Brandimarte L, Acquisti A, Loewenstein G. Misplaced confidences. Soc Psychol Personal Sci 2012 Aug 09;4(3):340-347. [doi: 10.1177/1948550612455931]
46. Yi J, Yuan G, Yoo C. The effect of the perceived risk on the adoption of the sharing economy in the tourism industry: the case of Airbnb. Inf Process Manag 2020 Jan;57(1):102108. [doi: 10.1016/j.ipm.2019.102108]
47. Hansen JM, Saridakis G, Benson V. Risk, trust, and the interaction of perceived ease of use and behavioral control in predicting consumers' use of social media for transactions. Comput Hum Behav 2018 Mar;80:197-206. [doi: 10.1016/j.chb.2017.11.010]
48. Hilverda F, Kuttschreuter M. Online information sharing about risks: the case of organic food. Risk Anal 2018 Sep 23;38(9):1904-1920. [doi: 10.1111/risa.12980] [Medline: 29570833]
49. Morris J. First Look: Internet Use in 2015. 2016. URL: https://www.ntia.doc.gov/blog/2016/first-look-internet-use-2015 [accessed 2020-04-20]
50. Lee Y, Queenie Li J. The value of internal communication in enhancing employees' health information disclosure intentions in the workplace. Public Relat Rev 2020 Mar;46(1):101872. [doi: 10.1016/j.pubrev.2019.101872]
51. Wang J, Huang J, Cheung CSK, Wong WN, Cheung NT, Wong MC. Adoption of an electronic patient record sharing pilot project: cross-sectional survey. J Med Internet Res 2020 Apr 06;22(4):e13761 [FREE Full text] [doi: 10.2196/13761] [Medline: 32250279]
52. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Inf Manag 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003]
53. Morris K, Yamamoto G, Sugiyama O, Luciano S, Tsutsumi T, Ohtsuki R, et al. Designing a mobile patient information sharing system using patients community members: perceptions of emergency physicians. Eur J Biomed Inf 2020;16(3):1-9 [FREE Full text]
54. Pavlou PA, Fygenson M. Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. MIS Quarterly 2006;30(1):115. [doi: 10.2307/25148720]
55. Lin W, Zhang X, Song H, Omori K. Health information seeking in the Web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Hum Behav 2016 Mar;56:289-294. [doi: 10.1016/j.chb.2015.11.055]
56. Gefen D, Karahanna E, Straub DW. Trust and TAM in online shopping: an integrated model. MIS Quarterly 2003;27(1):51. [doi: 10.2307/30036519]
57. Bartek MA, Truitt AR, Widmer-Rodriguez S, Tuia J, Bauer ZA, Comstock BA, et al. The promise and pitfalls of using crowdsourcing in research prioritization for back pain: cross-sectional surveys. J Med Internet Res 2017 Oct 06;19(10):e341 [FREE Full text] [doi: 10.2196/jmir.8821] [Medline: 28986339]

58. Pew Research Center. Internet/Broadband Fact Sheet. 2018. URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2021-04-25]
59. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
60. Hair J. Multivariate Data Analysis (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall; 2006:1-703.
61. Rise J, Kovac V, Kraft P, Moan IS. Predicting the intention to quit smoking and quitting behaviour: extending the theory of planned behaviour. Br J Health Psychol 2008 May;13(Pt 2):291-310. [doi: 10.1348/135910707X187245] [Medline: 17535498]

Abbreviations

CFI: comparative fit index
CPM: communication privacy management
CR: composite reliability
IoT: internet of things
MTurk: Mechanical Turk
RMSEA: root-mean-square error of approximation
SEM: structural equation modeling
TLI: Tucker-Lewis index
VIF: variance inflation factor

Edited by G Eysenbach; submitted 30.07.20; peer-reviewed by AS Alaqra, C Friedrich, S Uwizeyemungu; comments to author 24.09.20; revised version received 03.02.21; accepted 18.09.21; published 16.11.21.

Please cite as:
Abdelhamid M
Fitness Tracker Information and Privacy Management: Empirical Study
J Med Internet Res 2021;23(11):e23059
URL: https://www.jmir.org/2021/11/e23059
doi:10.2196/23059
PMID:34783672

Original Paper

Social Media Sharing of Articles About Measles in a European Context: Text Analysis Study

Dominik Wawrzuta[1], MD; Mariusz Jaworski[1], PhD; Joanna Gotlib[1], PhD; Mariusz Panczyk[1], PhD

Department of Education and Research in Health Sciences, Medical University of Warsaw, Warsaw, Poland

Corresponding Author:
Dominik Wawrzuta, MD
Department of Education and Research in Health Sciences
Medical University of Warsaw
Żwirki i Wigury 81
Warsaw, 02-091
Poland
Phone: 48 22 57 20 490
Fax: 48 22 57 20 491
Email: dwawrzuta@wum.edu.pl

Abstract

Background: Despite the existence of an effective vaccine, measles still threatens the health and lives of many Europeans. Notably, during the COVID-19 pandemic, measles vaccine uptake declined; as a result, after the pandemic, European countries will have to increase vaccination rates to restore the extent of vaccination coverage among the population. Because information obtained from social media are one of the main causes of vaccine hesitancy, knowledge of the nature of information pertaining to measles that is shared on social media may help create educational campaigns.

Objective: In this study, we aim to define the characteristics of European news about measles shared on social media platforms (ie, Facebook, Twitter, and Pinterest) from 2017 to 2019.

Methods: We downloaded and translated (into English) 10,305 articles on measles published in European Union countries. Using latent Dirichlet allocation, we identified main topics and estimated the sentiments expressed in these articles. Furthermore, we used linear regression to determine factors related to the number of times a given article was shared on social media.

Results: We found that, in most European social media posts, measles is only discussed in the context of local European events. Articles containing educational information and describing world outbreaks appeared less frequently. The most common emotions identified from the study's news data set were fear and trust. Yet, it was found that readers were more likely to share information on educational topics and the situation in Germany, Ukraine, Italy, and Samoa. A high amount of anger, joy, and sadness expressed within the text was also associated with a higher number of shares.

Conclusions: We identified which features of news articles were related to increased social media shares. We found that social media users prefer sharing educational news to sharing informational news. Appropriate emotional content can also increase the willingness of social media users to share an article. Effective media content that promotes measles vaccinations should contain educational or scientific information, as well as specific emotions (such as anger, joy, or sadness). Articles with this type of content may offer the best chance of disseminating vital messages to a broad social media audience.

(J Med Internet Res 2021;23(11):e30150) doi:10.2196/30150

KEYWORDS

measles; Facebook; Twitter; Pinterest; social media; vaccine; infodemiology; public health

Introduction

Background

The first measles vaccine was approved in 1963. Before the invention of this vaccine, measles caused 6 million deaths annually [1]. Since use of this vaccine became common, the number of measles cases worldwide has started to decline. A few years after the first vaccination campaign, the number of new measles cases plunged to one-twentieth of the previous figure [2]. In some countries, the cases of measles have been eliminated almost completely, although local epidemics still occur from time to time [3].

Despite the proven effectiveness of vaccination in saving millions of lives annually, some individuals have questioned the safety and long-term benefits of vaccine use. In 1998, Andrew Wakefield, a British physician and academic, published an article connecting the measles, mumps, and rubella (MMR) vaccine to autism among children [4]. This paper was retracted 12 years later; however, antivaccine activists still argue against vaccination. Antivaccination movements are one of the main obstacles for public health professionals in conducting vaccination campaigns. Vaccine hesitancy is a significant problem as only a high measles vaccination coverage—of about 95%—can enable the complete eradication of this disease. Currently, global vaccination coverage against measles is approximately 70%. Measles vaccines protect not only human lives but also the economies of low- and middle-income countries, generating US $58 for an investment of US $1 [5].

One example of the result of the decline in measles vaccination coverage was the outbreak in Samoa. In September 2019, in Samoa, a country with a small population of 200,000, a measles outbreak led to over 5000 infections and 83 deaths [6]. The reason behind the outbreak was the suspension of the country's immunization program by the Samoan government in July 2018 following the death of 2 children as a result of nurses' inadvertent use of curare muscle relaxant anesthetic instead of water to dilute the MMR vaccine. This led to a decrease in measles vaccination coverage in Samoa to 31% by the end of 2019 [7]. After multiple measles-related deaths, the authorities decided to organize a vaccination campaign. Approximately 95% of eligible people in Samoa were vaccinated against measles, which put an end to the outbreak [8]. This case shows how quickly the measles virus can spread when a vaccination program is suspended. Therefore, it is important to constantly monitor measles epidemiology and people's attitudes toward it to promptly prevent vaccine hesitancy.

This problem surrounding vaccination coverage has also been observed in Europe. The European Centre for Disease Prevention and Control (ECDC) data suggests that the second dose of MMR vaccine coverage is over 95% in only 5 European Union countries. This extent of coverage can ensure the immunity of the population against this disease and eliminate the chance of an outbreak [9]. In the European Union, the number of measles cases declined in 2020—mainly caused by the COVID-19 pandemic and the result of wearing masks, practicing social distancing, and conducting social lockdowns [10]—except for Romania and Bulgaria. In 2020, there were over 20 measles cases per million inhabitants in these 2 countries [11]. After the COVID-19 pandemic ends, European societies will probably notice an increase in measles morbidity as there was a decline in measles vaccine administration among children during the lockdowns [12]. Moreover, Europeans have been using social media more frequently during the COVID-19 pandemic. This may result in the growth of negative attitudes in the public toward vaccination as exposure to disinformation on social media increases vaccine hesitancy [13,14]. As social media contribute significantly, analyzing social media content and the activities of users can help better understand public attitudes and opinions regarding measles. The knowledge gathered from this analysis will inform the actions to increase MMR vaccine coverage and prevent the spread of misinformation.

Study Aim

We aim to characterize European measles news reports shared on social media platforms (ie, Facebook, Twitter, and Pinterest) during 2017 to 2019 (ie, the pre–COVID-19 period). For this purpose, we formulated the following 3 research questions: (1) What are the main topics of the articles on measles published in the European Union countries? (2) What sentiments are associated with these news articles? (3) Which features of the articles are associated with an increased number of shares on social media?

Methods

Data Collection and Preparation

We collected articles on measles that were shared on Facebook, Twitter, or Pinterest from 28 European Union countries.

First, we translated the word "measles" into all 23 official languages of the European Union using Wiktionary [15] and used a social media data analysis platform (ahrefs, Ahrefs Pte. Ltd [16]), which continuously collects, processes, and stores information from social media platforms about users' content and activities, to collect news articles. For each European Union country, we downloaded all articles on measles published from January 1, 2017, to December 31, 2019. We chose this time range because there was a significant increase in the number of measles infections in Europe in 2017 that lasted until the start of the COVID-19 pandemic [17]. We selected articles containing the word "measles" in the national language and published on websites with national domains (eg, ".de" domain for Germany, ".pl" for Poland). Our data set contained the URL of the article, the publication date, and the number of shares (ie, the total number of shares for all instances of the article) on Facebook, Twitter, and Pinterest.

As the next step, we obtained the full text of the articles (n=12,638) and read the content. To accomplish this task, we used a Python newspaper package (version 0.3.0) [18] that allows an automated download of the website content.

We automatically translated all non-English articles into English using Yandex Translate [19]. Finally, we removed all duplicate articles and those that had been improperly downloaded or translated. Our final data set comprised 10,305 articles. This dataset, containing the text of translated news, country of origin, and the total number of shares, was published and publicly made available on the Zenodo platform [20]. Finally, we processed the data in order to be able to apply the latent Dirichlet allocation (LDA) method. We used R packages (tidytext [21] and textstem [22]) to tokenize text; remove numbers, punctuation, and English stop words; and lemmatize all words.

Statistical Analysis

Topic Modeling

We used LDA [23] to identify the main topics of the 10,305 articles in our data set. We trained multiple LDA models with a different number of topics (ranging from 1 to 40). We then

analyzed perplexity and coherence levels to select the model that best describes our data set. In the next step, 2 researchers individually labeled the topics chosen by LDA to categorize them. The researchers analyzed not only the keywords assigned to each topic but also the content of the 20 articles with the highest amount of contribution to the topic. Initially, they independently described each topic with a freely chosen category. Then, they analyzed the created categories (without knowing the topic they were assigned to) and together created a unified set of categories (eg, education, Europe, and the world). Finally, they classified the topics again with a new set of categories. In this final stage, there were no discrepancies in assessment.

Sentiment Analysis

We calculated the main emotions associated with each article using the syuzhet R package [24]. This package uses the National Research Council Canada (NRC) Word-Emotion Association lexicon. The NRC lexicon is a set of 14,182 English words that are not just concerned with polarity (reporting positive or negative words) but associated with 8 fundamental emotions introduced by Plutchik [25] (ie, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) [26]. These words were labeled manually by crowdsourcing—each word could be associated with more than one emotion). The sentiment of an article is the sum of emotions related to the words that make it up. The occurrence of each word from one of the categories in the article translates as “1” in the sentiment score for that category. Finally, each article is scored for each sentiment category [27].

Linear Regression

We considered a linear regression model to find the factors determining the number of shares of an article on social media (ie, the dependent variable). We used forward selection regression to create our model. We also used Cook distance method to identify and remove outliers [28] and variance inflation factor to check the existence of collinearity [29]. As independent variables, we used topics generated by the LDA model, the emotions related to the articles, the number of new national measles cases in the month when an article was published, the population of the country, and the percentage of active social media users in each country. The number of new monthly measles cases for each country was collected from the ECDC website [30]. Data on the populations of European countries were obtained from the Eurostat database [31], and the proportion data of active social media users in each country were acquired from the Statista website [32].

Results

Sample Description

After article selection and data processing, we had a final sample of 10,305 measles-related articles, published between January 1, 2017, and December 31, 2019, in European Union countries. The highest number of published articles retrieved was from Italy, but the articles that were the most shared ones were from the United Kingdom. Table 1 shows the number of articles from each country, and the sum and average number of shares. In Table S1 of Multimedia Appendix 1, we present more detailed yearly aggregated data.

Table 1. Description of collected data (N=10,305).

Country	Articles, n (%)	Shares	
		Total	Mean per article (SD)
Austria	200 (1.94)	34,539	173 (425)
Belgium	250 (2.43)	89,442	358 (1095)
Bulgaria	404 (3.92)	23,418	58 (346)
Croatia	53 (0.51)	8395	158 (281)
Republic of Cyprus	9 (0.09)	250	28 (52)
Czech Republic	51 (0.49)	78,043	1530 (3154)
Denmark	114 (1.11)	92,603	812 (2274)
Estonia	13 (0.13)	2546	196 (349)
Finland	119 (1.15)	86,076	723 (2136)
France	1252 (12.15)	633,111	506 (1826)
Germany	1132 (10.98)	914,665	808 (4420)
Greece	697 (6.76)	57,082	82 (792)
Hungary	116 (1.13)	28,851	249 (568)
Ireland	167 (1.62)	32,226	193 (389)
Italy	2025 (19.65)	1,462,172	722 (2983)
Latvia	3 (0.03)	927	309 (520)
Lithuania	1 (0.01)	156	156 (0)
Luxembourg	20 (0.19)	2609	130 (212)
Malta	0 (0)	0	N/A[a]
Netherlands	313 (3.04)	68,705	220 (607)
Poland	216 (2.10)	46,658	216 (1042)
Portugal	964 (9.35)	200,812	208 (1116)
Romania	544 (5.28)	44,363	82 (459)
Slovakia	21 (0.20)	4300	205 (255)
Slovenia	25 (0.24)	2450	98 (232)
Spain	664 (6.44)	954,613	1438 (18,634)
Sweden	253 (2.46)	86,958	344 (1361)
United Kingdom	679 (6.59)	2,012,118	2963 (23,726)

[a]N/A: not applicable.

Topic Modeling

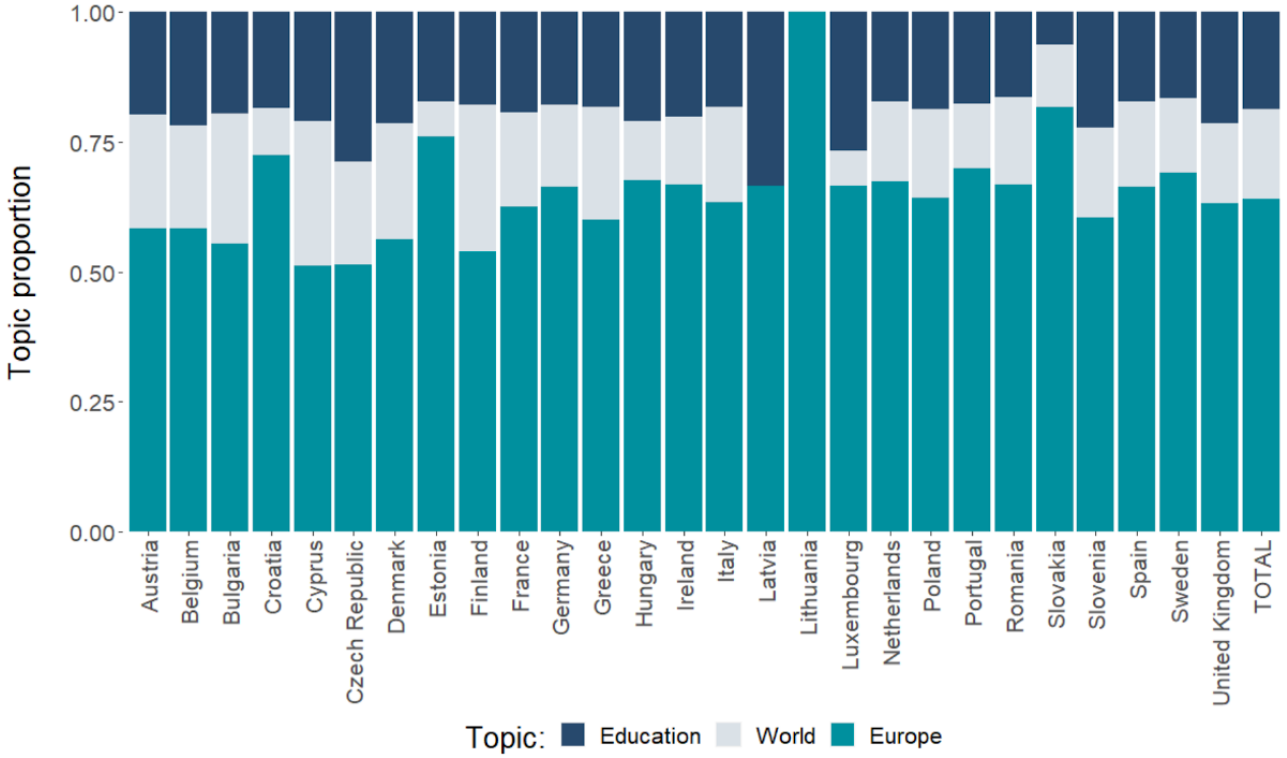
We found that 13 is the best number of topics to describe all collected news articles, accounting for the perplexity and coherence values (Table S2 of Multimedia Appendix 1). After labeling, we discovered that these topics can be aggregated into 3 clusters. The first group contains the topics featured in educational articles, which describe the signs and symptoms of measles infection, debunk antivaccine claims, and explain scientific advancements in the prevention of this disease. The topics in the second group are related to European information, which contain country-specific information on measles outbreaks and health policies. The last group includes topics related to countries outside Europe, including measles cases in the United States, and the Samoa measles outbreak. Table 2 presents the main words connected with specific topics and their classification to general groups.

Table 2. Topics, their classification, and key words.

Topic	Main words	Group classification
Topic 1	immune, immune_system, study, cell, system, researcher, cancer, antibody, virus, memory	Education
Topic 2	child, measles, hospital, Samoa, epidemic, Sweden, people, disease, campaign, patient	World
Topic 3	county, York, measles, city, orthodox, Jewish, disease, USA, Brooklyn, confirm	World
Topic 4	measles, rash, symptom, fever, infection, disease, day, infect, cough, virus	Education
Topic 5	vaccine, parent, autism, child, ani, polio, Wakefield, diphtheria, study, vaccinate_child	Education
Topic 6	MMR, GP, December, England, measles, UK, dose_MMR, jab, HSE, NHS	Europe
Topic 7	vaccination, school, federal, Spahn, measles_vaccination, Germany, day_care, measles, CDU, mandatory	Europe
Topic 8	DGS, health, confirm, directorate, measles, outbreak, age, Portugal, UK, Lisbon	Europe
Topic 9	hospital, measles, Italy, health, Catania, ship, vaccinate, Sicily, hospitalize, region	Europe
Topic 10	country, world, Europe, organization, world_health, health_organization, European, measles, increase, Ukraine	Europe
Topic 11	measles, Roma, CDC, health, dose, Greece, vaccine, population, diseases, Spain	Europe
Topic 12	France, measles, health, agency, Aquitaine, epidemic, health_France, vaccinate, Poitiers, people	Europe
Topic 13	obligation, dolphin, Cicciobello, decree, sport, doll, Burioni, certification, time, market	Europe

Using these 3 clustered meta-topics, we evaluated the popularity of all topics in each country. Figure 1 shows that all of the European countries mainly write about the topic in the European context. Educational topics are more popular than world topics, although the difference between the two varies between countries. In Multimedia Appendix 1, we present sample articles that were highly connected to specific topics.

Figure 1. Distribution of topics across all countries analzyed.



Sentiment Analysis

We analyzed what emotions are connected to measles-related articles in Europe. Figure 2 shows the frequency of appearance of the words associated with a specific emotion in our data set. The most common emotions were fear and trust.

We also determined which words contributed the most to the emotion levels in our data set. For each emotion, we reviewed the 15 most popular words from our data set, as shown in Table 3.

Figure 2. Number of words with a certain sentiment in the data set of articles analyzed.

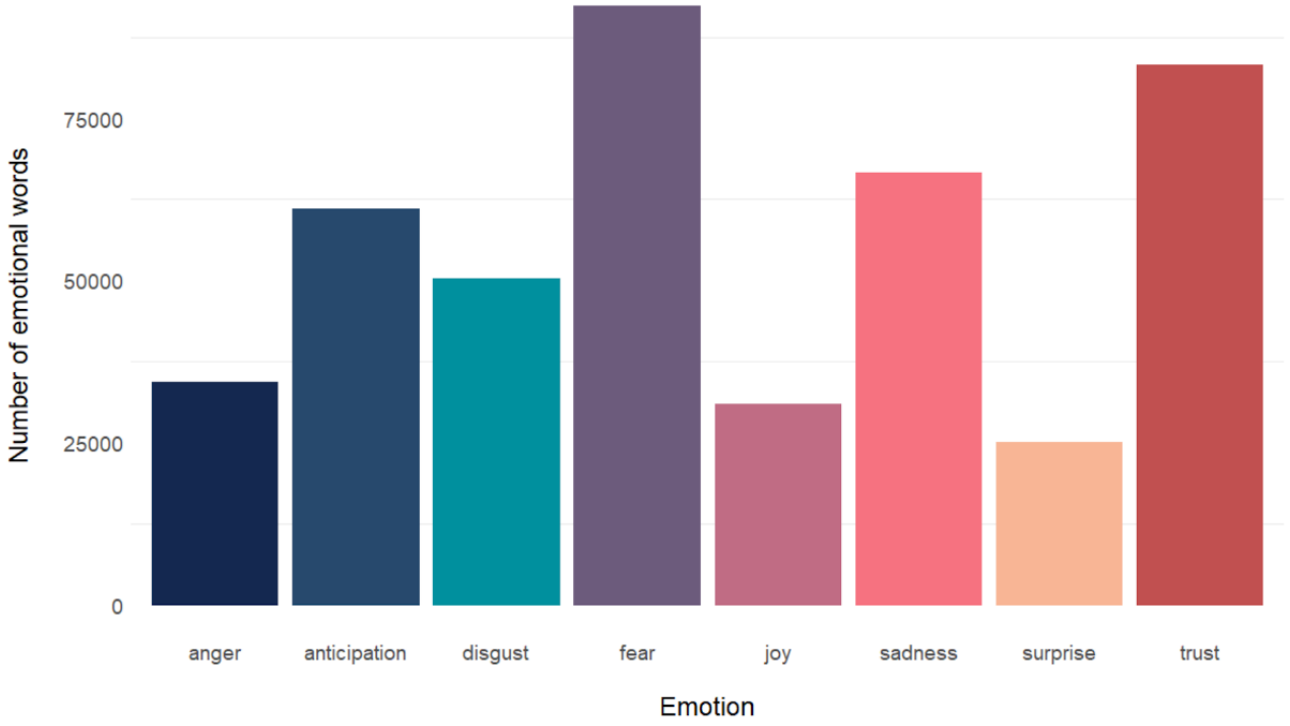


Table 3. Emotions and connected words.

Emotion	Connected words
Fear	disease, infection, death, hospital, epidemic, risk, infectious, contagious, die, medical, fever, prevent, bear, warn, emergency
Trust	accord, hospital, school, doctor, medical, immunization, county, system, organization, authority, recommend, ministry, continue, level, director
Sadness	disease, death, hospital, epidemic, infectious, die, late, sick, emergency, illness, leave, mother, fatal, fall, bad
anticipation	child, death, time, patient, epidemic, risk, public, medical, organization, result, continue, warn, start, prevention, develop
Disgust	disease, death, epidemic, infect, infectious, contagious, rash, cough, sick, bad, ill, nose, elimination, lose, treat
Anger	disease, death, epidemic, bear, fatal, eradicate, bad, ill, force, fear, elimination, victim, lose, treat, fight
Joy	child, organization, ministry, baby, majority, infant, mother, childhood, achieve, safe, grow, treat, create, save, progress
Surprise	death, epidemic, organization, warn, emergency, infant, leave, alarm, expect, catch, lose, treat, sneeze, break, vote

Evidently, fear is connected to words describing the harmfulness of measles. Trust is associated with the words related to health care systems and protection. *Sadness* is conveyed when describing the adverse effects of the disease; *anticipation* is conveyed when talking about the actions being taken to reduce the spread of the disease; *disgust* is associated with the characteristics and signs of measles; *anger* is about the fight against measles; *joy* is about children and protecting them from the disease; and *surprise* is about emergent events. In Multimedia Appendix 1, we present sample articles highly connected to specific emotions.

Linear Regression

We used linear regression to identify the variables related to the number of shares of an article on social media. Table 4 presents the results of the forward selection regression. In the multicollinearity analysis, all variance inflation factor scores were determined to be lower than 5. Therefore, we assumed that there is no collinearity among variables. Detailed results are presented in Multimedia Appendix 1.

We discovered that readers were more likely to share information on topics 1, 4, and 5, which convey educational information. Topics 2, 7, 10, and 13 describe the situation in Samoa, Germany, Ukraine, and Italy, and these topics were also positively associated with the number of total shares. Topic 6, which describes the situation in the United Kingdom, resulted in a reduced number of shares. Topic 4, which describes the signs and symptoms of measles, had the highest impact on the average of shareability.

With regard to emotions, on average, a higher proportion of anger, joy, and sadness expressed in an article was associated with a higher number of article shares, whereas surprise in the article was associated with reduced number of shares. Moreover,

articles published in countries with a higher population and a higher number of active social media users understandably received more shares on social media.

Table 4. Regression results (adjusted R^2=0.04852).

Variable	Beta	Standard error	*t* value (*df*)[a]	*P* value
Constant	–1994	282	–7.08	<.001
Topic 1	1025	218	4.695	<.001
Topic 2	711	127	5.615	<.001
Topic 4	1833	145	12.651	<.001
Topic 5	475	196	2.426	.015
Topic 6	–512	171	–2.994	.003
Topic 7	1024	122	8.42	<.001
Topic 10	616	114	5.388	<.001
Topic 12	184	124	1.487	.14
Topic 13	375	181	2.072	.04
Anger	54	15	3.512	<.001
Joy	53	14	3.66	<.001
Surprise	–50	20	–2.551	.01
Sadness	38	13	2.828	.005
Disgust	–24	15	–1.603	.11
Social media users	22	4	5.112	<.001
Population	0.000018	0.000001	9.393	<.001

[a]*df*=10,254.

Discussion

Sample Description

The number of articles published from 2017 to 2019 varies across the different countries included in this study. The highest number of articles was published in Italy, which is likely because of the high number of measles cases reported in Italy. During this period, a total of 9252 measles cases were reported in Italy [30]—the highest reported in the European Union. The articles published in the United Kingdom received the greatest number of shares, which might be attributed to the popularity of the English language worldwide. In our data set, less than one relevant article per month was published in Malta, Slovenia, Slovakia, Lithuania, Estonia, Latvia, and Luxembourg. These countries have small populations, and most reported a low number of measles cases during 2017 to 2019. Latvia had a relatively higher number of measles cases, but as was observed (Table 1), despite the low number of measles-related articles published, they were extensively shared on social media. Topic Modeling

European media mostly published news about local events, reporting on almost all significant outbreaks of measles in Europe. However, several of the events received special media attention. The decision of the German government to make the measles vaccine mandatory in response to an increase in measles cases has been frequently discussed in those articles [33]. The measles outbreak in the Aquitaine region in France, caused by insufficient vaccination coverage, is also one of the most described events in the media [34]. Furthermore, measles clusters in the United Kingdom, Portugal, Italy, and Greece are also frequently mentioned in these articles. Some of these articles describe the case of the Cicciobello doll in Italy. The Cicciobello doll is a toy that pretends to be suffering from measles and that children can cure with plasters or cream. Experts have criticized this doll for banalizing such a severe disease [35].

The theme of world news is dominated by 2 events: One is the measles outbreak in the Orthodox Jewish community in New York [36], and the second is the measles outbreak in Samoa [7]. Both these events are interesting because they are concerned with relatively small outbreaks. During the years of analysis, millions of people contracted measles in Africa, Asia, and South America, but this did not attract the attention of European media. As indicated by our data set, media attention was mainly drawn to outbreaks in small, specific communities, and not necessarily to events that had the greatest impact on the lives of millions of people.

Educational themes focus mainly on 3 threads. Some articles describe the symptoms of measles, reflecting the readers' interest in the signs, symptoms, and causes of the disease and their desire to recognize them. The second topic is related to the scientific findings on measles, including studies on the potential oncolytic activity of this virus [37]. The last educational topic dispels the doubts—raised by Wakefield's paper, which has since been retracted—related to the risk of autism allegedly caused by the

MMR vaccine [4]. Moreover, another paper published in 2019, of a nationwide cohort study conducted in Denmark that found that MMR vaccination does not increase the risk for autism, was widely discussed in the media [38].

Sentiment Analysis

In the course of human life, up to 34,000 different emotions can be distinguished [39]. Psychologist Robert Plutchik proposed the theory of 8 basic emotions, which have developed evolutionarily and are innate in humans and help them survive [25,40]. As a result of combining these basic emotions, more complex emotions responsible for specific experiences arise (eg, joy + trust = love; trust + fear = submission). Basic emotions are triggered by specific events, eliciting an appropriate response that has an evolutionary function. Table 5 presents the characteristics of these basic emotions as proposed by Plutchik [40].

Table 5. Characteristics of basic emotions.

Emotion	Stimulus event	Behavioral reaction	Function	Opposite emotion
Joy	Gain of a valued object	Retain or repeat	Gain resources	Sadness
Trust	Member of the group	Groom	Mutual support	Disgust
Fear	Threat	Escape	Safety	Anger
Surprise	Unexpected event	Stop	Gain time to orient	Anticipation
Sadness	Loss of a valued object	Cry	Reattach to the lost object	Joy
Disgust	Unpalatable object	Vomit	Eject poison	Trust
Anger	Obstacle	Attack	Destroy obstacle	Fear
Anticipation	New territory	Map	Knowledge of territory	Surprise

Emotions are an important part of media articles, as emotional stories attract readers' attention [41]. The most common emotion identified from all collected articles was *fear*. The creation of fear-based articles is consistent with the results of studies that show that fear-related news reports automatically attract public attention [42]. In the case of the articles analyzed in this study, the emotion *fear* was mainly concerned with the fear of the disease and its consequences. This emotion is so strong that sometimes the fear of disease is worse for a patient than the disease itself [43].

The analyzed articles were also highly connected with *trust*. There is a natural link between fear and trust because it is in response to concerns about a dangerous disease that we place our trust in the health care system and the vaccines that protect us against the pathogen. Research conducted during the Ebola epidemic showed that trust in the health care system increased during the outbreak [44].

In the field of marketing, the relationship between the basic emotions in the text and its potential to go viral has often been examined. An analysis of 7000 articles published in The New York Times revealed that positive content tends to go viral more than negative content. More specifically, this study has demonstrated that sad content is less likely to go viral, whereas articles expressing anger or anxiety result in higher number of shares on social media [45]. Teixeira [46] analyzed thousands of reactions to several advertisements and found that maintaining viewers' engagement levels is associated with the emotions of joy and surprise. Libert and Tynski [47] also found that emotional activation is the key to viral success. They used Plutchik's set of emotions [25] and found that negative emotions are not commonly present in highly viral content. Surprise and anticipation are also extremely common in highly viral content.

Linear Regression

We identified a few features of news articles that are associated with an increased number of shares on social media. All educational topics are positively connected with the number of shares. This shows that social media users generally prefer to share general educational news over informational ones. This finding is consistent with the results of previous research, wherein it was proved that Facebook users are more likely to share "soft" news related to children, health, and education than "hard" news related to politics or urgent occurrences [48]. Sharing educational and scientific articles can also be associated with the willingness to increase one's credibility, self-confidence, and self-esteem from other social media users [49].

The results also show that information on the events in Sweden, Germany, Italy, and Ukraine was shared frequently, whereas the publication of information on the situation in the United Kingdom was negatively associated with the number of social media shares. The interest in measles in Ukraine may have been generated as a result of the country having the highest number of measles cases in Europe. In 2017-2019, the number of measles cases in Ukraine was around 100,000 [50]. This situation was similar to the interest in measles cases in Italy, where the highest number of cases among all countries in the European Union was reported during the study period [30]. Furthermore, research conducted by Facebook showed that country-based or cultural differences have an impact on Facebook activity. In the United Kingdom, social media users are younger but less active than those in Germany or Sweden [51].

The findings of previous research analyzing what type of web-based content become viral are generally in line with our results. They indicate that articles containing positive emotions or anger are more likely to be shared [45].

Conclusions

Articles on measles shared on social media in Europe primarily report on European events, and only a small proportion of articles report on educational news or international measles-related events. The international events mainly describe outbreaks that have occurred in a small number of infected people but are interesting from an epidemiological point of view. The distribution of topics covered by the media is similar across all European Union countries.

In this study, the two main emotions expressed in the analyzed measles-related articles were fear and trust. These emotions appeared in the articles most frequently. However, these emotions were not associated with frequent sharing of articles on social media. We found that an article has a high probability to drive public discourse if it contains educational or scientific information, as well as specific emotions (ie, anger, joy, or sadness). Making media content based on these principles can facilitate the creation of effective messages against measles vaccine hesitancy. Articles that follow these principles offer the best chance of disseminating information to a broad audience on social media and influencing the mindset of the public regarding vaccines.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables providing details of the analyses conducted.
[DOC File , 231 KB - jmir_v23i11e30150_app1.doc]

References

1. Goodson JL, Seward JF. Measles 50 years after use of measles vaccine. Infect Dis Clin North Am 2015 Dec;29(4):725-743. [doi: 10.1016/j.idc.2015.08.001] [Medline: 26610423]
2. Conis E. Measles and the modern history of vaccination. Public Health Rep 2019 Feb 14;134(2):118-125 [FREE Full text] [doi: 10.1177/0033354919826558] [Medline: 30763141]
3. Patel MK, Dumolard L, Nedelec Y, Sodha SV, Steulet C, Gacic-Dobo M, et al. Progress toward regional measles elimination - Worldwide, 2000-2018. MMWR Morb Mortal Wkly Rep 2019 Dec 06;68(48):1105-1111 [FREE Full text] [doi: 10.15585/mmwr.mm6848a1] [Medline: 31805033]
4. Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. BMJ 2011 Jan 05;342(jan05 1):c7452-c7452. [doi: 10.1136/bmj.c7452] [Medline: 21209060]
5. Durrheim DN. Measles eradication—retreating is not an option. Lancet Infect Dis 2020 Jun;20(6):e138-e141. [doi: 10.1016/s1473-3099(20)30052-9]
6. Craig AT, Heywood AE, Worth H. Measles epidemic in Samoa and other Pacific islands. The Lancet Infectious Diseases 2020 Mar;20(3):273-275. [doi: 10.1016/s1473-3099(20)30053-0]
7. Isaacs D. Lessons from the tragic measles outbreak in Samoa. J Paediatr Child Health 2020 Jan 15;56(1):175-175. [doi: 10.1111/jpc.14752] [Medline: 31943476]
8. WHO. Measles Outbreak in The Pacific - Situation Report No. 9. Joint WHO/UNICEF Measles Outbreak Response and Preparedness in the Pacific. 2019. URL: https://www.who.int/docs/default-source/wpro---documents/dps/outbreaks-and-emergencies/measles-2019/measles-pacific-who-unicef-sitrep-20200103.pdf [accessed 2021-10-21]
9. Vaccination coverage for the second dose of measles-containing vaccine, EU/EEA, 2018. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/publications-data/vaccination-coverage-second-dose-measles-containing-vaccine-eueea-2018 [accessed 2021-10-21]
10. Wu D, Liu Q, Wu T, Wang D, Lu J. The impact of COVID-19 control measures on the morbidity of varicella, herpes zoster, rubella and measles in Guangzhou, China. Immun Inflamm Dis 2020 Dec 07;8(4):844-846 [FREE Full text] [doi: 10.1002/iid3.352] [Medline: 33026189]
11. Measles notification rate per million population by country, January 2020 - December 2020. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/publications-data/measles-notification-rate-million-population-country-january-2020-december-2020 [accessed 2021-10-21]
12. Santoli JM, Lindley MC, DeSilva MB, Kharbanda EO, Daley MF, Galloway L, et al. Effects of the COVID-19 pandemic on routine pediatric vaccine ordering and administration - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 May 15;69(19):591-593 [FREE Full text] [doi: 10.15585/mmwr.mm6919e2] [Medline: 32407298]
13. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct 23;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
14. Wawrzuta D, Jaworski M, Gotlib J, Panczyk M. Characteristics of antivaccine messages on social media: systematic review. J Med Internet Res 2021 Jun 04;23(6):e24564 [FREE Full text] [doi: 10.2196/24564] [Medline: 34085943]
15. Wiktionary. URL: https://www.wiktionary.org/ [accessed 2021-10-21]
16. Ahrefs - SEO Tools & Resources to Grow Your Search Traffic. URL: https://ahrefs.com/ [accessed 2021-10-21]

17. Number of cases of measles reported monthly in the European Economic Area (EEA) from 1999 to 2021. Statista. URL: https://www.statista.com/statistics/1122761/measles-cases-in-europe-timeline/ [accessed 2021-10-21]
18. Ou-Yang L. News, full-text, and article metadata extraction in Python 3. GitHub. URL: https://github.com/codelucas/newspaper [accessed 2021-10-21]
19. Yandex Translate. yandex.com. URL: https://translate.yandex.com/ [accessed 2021-10-21]
20. Wawrzuta D. European articles about measles. Dataset on Zenodo. URL: https://zenodo.org/record/4725249 [accessed 2021-10-21]
21. Silge J, Robinson D. tidytext: text mining and analysis using tidy data principles in R. JOSS 2016 Jul;1(3):37. [doi: 10.21105/joss.00037]
22. Rinker T. Tools for fast text stemming & lemmatization. URL: http://github.com/trinker/textstem [accessed 2021-10-21]
23. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text]
24. Jockers M. Introduction to the syuzhet package. URL: https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html [accessed 2021-10-21]
25. Plutchik R. A general psychoevolutionary theory of emotion. In: Emotion: Theory, research, and experience: Vol. 1. Theories of emotion. New York: Academic Press; 1980.
26. Mohammad SM, Turney PD. NRC Emotion Lexicon. URL: https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm [accessed 2021-10-21]
27. Naldi M. A review of sentiment computation methods with R packages. arXiv 2019:1-11.
28. Stevens JP. Outliers and influential data points in regression analysis. Psychol Bull 1984;95(2):334-344. [doi: 10.1037/0033-2909.95.2.334]
29. Thompson CG, Kim RS, Aloe AM, Becker BJ. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl Soc Psych 2017 Feb;39(2):81-90. [doi: 10.1080/01973533.2016.1277529]
30. Measles. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/measles [accessed 2021-10-21]
31. Population on 1 January by age and sex. Eurostat. URL: https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en [accessed 2021-10-21]
32. Tankovska H. Active social media penetration in selected European countries in 2020. Statista. URL: https://www.statista.com/statistics/295660/active-social-media-penetration-in-european-countries/ [accessed 2021-10-21]
33. Connolly K. German parliament approves compulsory measles vaccinations. The Guardian. URL: https://www.theguardian.com/world/2019/nov/14/german-parliament-approves-compulsory-measles-vaccinations [accessed 2021-10-21]
34. Bernadou A, Astrugue C, Méchain M, Le Galliard V, Verdun-Esquer C, Dupuy F, et al. Measles outbreak linked to insufficient vaccination coverage in Nouvelle-Aquitaine region, France, October 2017 to July 2018. Eurosurveillance 2018;23(30):1-5. [doi: 10.2807/1560-7917.es.2018.23.30.1800373]
35. Monella L, Cuddy A. Italy's measles doll sparks controversy. Euronews. URL: https://www.euronews.com/2018/04/06/italy-s-measles-doll-sparks-controversy [accessed 2021-10-21]
36. McDonald R, Ruppert P, Souto M, Johns D, McKay K, Bessette N, et al. Notes from the field: measles outbreaks from imported cases in orthodox Jewish communities - New York and New Jersey, 2018-2019. MMWR Morb Mortal Wkly Rep 2019 May 17;68(19):444-445 [FREE Full text] [doi: 10.15585/mmwr.mm6819a4] [Medline: 31095533]
37. Russell L, Peng K. The emerging role of oncolytic virus therapy against cancer. Chin Clin Oncol 2018 Apr;7(2):16-16 [FREE Full text] [doi: 10.21037/cco.2018.04.04] [Medline: 29764161]
38. Hviid A, Hansen JV, Frisch M, Melbye M. Measles, mumps, rubella vaccination and autism. Ann Intern Med 2019 Mar 05;170(8):513. [doi: 10.7326/m18-2101]
39. Managing emotions. Deloitte.: Deloitte; 2020. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Images/Misc/infographic/managing-emotions-infographic.pdf [accessed 2021-10-20]
40. Plutchik R. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am Sci 2001;89(4):344-350. [doi: 10.1511/2001.4.344]
41. Bagozzi RP, Gopinath M, Nyer PU. The role of emotions in marketing. J Acad Mark Sci 1999 Apr 01;27(2):184-206. [doi: 10.1177/0092070399272005]
42. Schmidt LJ, Belopolsky AV, Theeuwes J. Potential threat attracts attention and interferes with voluntary saccades. Emotion 2015 Jun;15(3):329-338. [doi: 10.1037/emo0000041] [Medline: 25527964]
43. Singh T, Newby DE. Is the fear of disease worse than the disease itself? Heart 2021 Jan 24;107(2):91-92. [doi: 10.1136/heartjnl-2020-318207] [Medline: 33234671]
44. Nuriddin A, Jalloh M, Meyer E, Bunnell R, Bio F, Jalloh M, et al. Trust, fear, stigma and disruptions: community perceptions and experiences during periods of low but ongoing transmission of Ebola virus disease in Sierra Leone, 2015. BMJ Global Health 2018;3(2):e000410. [doi: 10.1136/bmjgh-2017-000410]
45. Berger J, Milkman K. What Makes Online Content Viral? J Mark Res 2012 Apr 01;49(2):192-205. [doi: 10.1509/jmr.10.0353]
46. Teixeira T. The New Science of Viral Ads. Harvard Business Review. 2012. URL: https://hbr.org/2012/03/the-new-science-of-viral-ads [accessed 2021-10-26]

47. Libert K, Tynski K. Research: The Emotions that Make Marketing Campaigns Go Viral. Harvard Business Review. 2013. URL: https://hbr.org/2013/10/research-the-emotions-that-make-marketing-campaigns-go-viral [accessed 2021-10-26]
48. Kalsnes B, Larsson AO. Understanding news sharing across social media. Journalism Studies 2017 Mar 23;19(11):1669-1688. [doi: 10.1080/1461670x.2017.1297686]
49. Lee CS, Ma L, Goh DHL. Why do people share news in social media? In: Zhong N, Callaghan V, Ghorbani AA, Hu B, editors. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2011:129-140.
50. Rodyna R. Measles situation in Ukraine during the period 2017-2019. European Journal of Public Health 2019;29:1. [doi: 10.1093/eurpub/ckz186.496]
51. Cheng J, Burke M, de Gant B. Country Differences in Social Comparison on Social Media. Proc ACM Hum Comput Interact 2021 Jan 05;4(CSCW3):1-26. [doi: 10.1145/3434179]

Abbreviations

ECDC: The European Centre for Disease Prevention and Control
LDA: latent Dirichlet allocation
MMR: measles, mumps, and rubella

Edited by R Kukafka; submitted 03.05.21; peer-reviewed by C Slavik, M Das; comments to author 22.05.21; revised version received 17.07.21; accepted 12.08.21; published 08.11.21.

Please cite as:
Wawrzuta D, Jaworski M, Gotlib J, Panczyk M
Social Media Sharing of Articles About Measles in a European Context: Text Analysis Study
J Med Internet Res 2021;23(11):e30150
URL: https://www.jmir.org/2021/11/e30150
doi:10.2196/30150
PMID:34570715

Original Paper

Reactance to Social Authority in a Sugar Reduction Informational Video: Web-Based Randomized Controlled Trial of 4013 Participants

Violetta Hachaturyan[1], MSc; Maya Adam[2], MD; Caterina Favaretti[1], MSc; Merlin Greuel[1], MGH; Jennifer Gates[3], BA; Till Bärnighausen[1,4,5], MD; Alain Vandormael[1], PhD

[1]Heidelberg Institute of Global Health, University of Heidelberg, Heidelberg, Germany

[2]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, United States

[3]Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Africa Health Research Institute, Durban, South Africa

[5]Harvard Center for Population and Development Studies, Cambridge, MA, United States

Corresponding Author:
Alain Vandormael, PhD
Heidelberg Institute of Global Health
University of Heidelberg
130.3 Im Neuenheimer Feld
Heidelberg, 69210
Germany
Phone: 49 6221 56 5344
Email: alain.vandormael@uni-heidelberg.de

Abstract

Background: Short and animated story-based (SAS) videos can be an effective strategy for promoting health messages. However, health promotion strategies often motivate the rejection of health messages, a phenomenon known as reactance. In this study, we examine whether the child narrator of a SAS video (perceived as nonthreatening, with low social authority) minimizes reactance to a health message about the consumption of added sugars.

Objective: This study aims to determine whether our SAS intervention video attenuates reactance to the sugar message when compared with a content placebo video (a health message about sunscreen) and a placebo video (a nonhealth message about earthquakes) and determine if the child narrator is more effective at reducing reactance to the sugar message when compared with the mother narrator (equivalent social authority to target audience) or family physician narrator (high social authority) of the same SAS video.

Methods: This is a web-based randomized controlled trial comparing an intervention video about sugar reduction narrated by a child, the child's mother, or the family physician with a content placebo video about sunscreen use and a placebo video about earthquakes. The primary end points are differences in the antecedents to reactance (proneness to reactance, threat level of the message), its components (anger and negative cognition), and outcomes (source appraisal and attitude). We performed analysis of variance on data collected (N=4013) from participants aged 18 to 59 years who speak English and reside in the United Kingdom.

Results: Between December 9 and December 11, 2020, we recruited 38.62% (1550/4013) men, 60.85% (2442/4013) women, and 0.52% (21/4013) others for our study. We found a strong causal relationship between the persuasiveness of the content promoted by the videos and the components of reactance. Compared with the placebo (mean 1.56, SD 0.63) and content placebo (mean 1.76, SD 0.69) videos, the intervention videos (mean 1.99, SD 0.83) aroused higher levels of reactance to the message content (P<.001). We found no evidence that the child narrator (mean 1.99, SD 0.87) attenuated reactance to the sugar reduction message when compared with the physician (mean 1.95, SD 0.79; P=.77) and mother (mean 2.03, SD 0.83; P=.93). In addition, the physician was perceived as more qualified, reliable, and having more expertise than the child (P<.001) and mother (P<.001) narrators.

Conclusions: Although children may be perceived as nonthreatening messengers, we found no evidence that a child narrator attenuated reactance to a SAS video about sugar consumption when compared with a physician. Furthermore, our intervention

videos, with well-intended goals toward audience health awareness, aroused higher levels of reactance when compared with the placebo videos. Our results highlight the challenges in developing effective interventions to promote persuasive health messages.

Trial Registration: German Clinical Trials Registry DRKS00022340; https://tinyurl.com/mr8dfena

International Registered Report Identifier (IRRID): RR2-10.2196/25343

(J Med Internet Res 2021;23(11):e29664) doi:10.2196/29664

KEYWORDS

sugar reduction; reactance; animated video; digital intervention; health communication

Introduction

Background

Digital health interventions that promote educational messages to improve knowledge and change behaviors are commonly used as effective health promotion strategies. Comparisons of digital behavior change interventions with traditional face-to-face interventions indicate that web-based health promotion is generally at least as effective as conventional approaches and has several advantages, such as low cost, feasibility, and scalability [1,2]. Existing evidence suggests that the use of pictures [3], entertainment education [4], digital storytelling, and narrative structured messages [4] are some of the successful approaches for creating compelling, evidence-based health messages. As narrative messages do not include a direct, controlling language and words, such as *should*, *must*, and *required* [5,6], and conceal the persuasive intent, they can be more effective when compared with traditional health communication strategies [7]. To further explore the effectiveness of these innovative strategies in health communication, we created a short and animated story-based (SAS) video that draws from entertainment-education media, communication theory, and the animated entertainment industry to promote healthy behaviors over social media channels [8]. However, SAS videos may face the same challenges faced by other traditional methods of health persuasion that often arouse a motivation to reject the health message, a phenomenon known as reactance [9].

The theory of reactance consists of 4 elements: (1) perceived freedom, which individuals possess insofar as they are aware of it and can enact it; (2) threat to freedom, when pressure is exerted that makes it difficult to enact that freedom; (3) reactance, which refers to the motivation to reestablish the threatened freedom; and (4) direct restoration, which involves the freedom of the individual to perform the forbidden behavior [5]. Reactance plays a critical role in determining the effectiveness and acceptance of health promotion interventions. This has led to an active research agenda in developing strategies to reduce reactance in several areas, such as e-cigarette use [10], littering [11], alcohol [12], and eating behaviors [13], among others [14-16].

Objectives

In this study, we produced a SAS video with a message about reducing the consumption of added sugars. Designed for a diverse and global audience, the animated video uses a narrative based structure to minimize reactance to the sugar message. For our first hypothesis, we assess if our narrative-based, animated video is effective at attenuating reactance to a persuasive health message.

> *We hypothesize that there is a causal pathway between exposure to a SAS video about sugar intake reduction and reactance, its antecedents, and outcomes.* [Hypothesis 1]

While designing the video, we considered the degree of social authority that should be assigned to the narrator. First, we selected the traditional role of a physician who has high social authority. Although health experts and physicians are often used to promote health messages [17-19], previous studies have shown that individuals may perceive these messengers as coercive, threatening, or having a hidden persuasive intent [20], which could sustain or heighten reactance [9,21]. Therefore, we considered a child narrator as a potentially powerful and effective narrator, as a child may be perceived as nonthreatening, neutral, and without having an ulterior motive. To date, we were unable to find prior research on the effectiveness of a child narrator to attenuate reactance using a narrative-based, animated video format. The second hypothesis is as follows:

> *We hypothesize that a SAS video about sugar consumption narrated by a child (low social authority) will arouse less reactance when compared with a video narrated by the child's mother (equivalent social authority to the target audience) or the family doctor (high social authority).* [Hypothesis 2]

We used a randomized controlled trial (RCT) to measure the causal effect of social authority on reactance to a short, animated video about sugar intake reduction. The randomization ensures that there are no systematic differences introduced at the enrollment stage, which may lead to potential bias. In addition, an innovative feature of our study is the use of 2 placebo groups, which enabled us to isolate the health awareness and content effects of the intervention video.

Methods

Trial Design

This study is a web-based RCT with 3 intervention arms (*arms 1-3*), a content placebo arm (*arm 4*), and a placebo arm (*arm 5*). The participants in each intervention arm watched the same sugar video narrated by a child (*arm 1:* low social authority), the child's mother (*arm 2:* equivalent social authority), or the physician (*arm 3:* high social authority). *Arm 4* watched a content placebo video with a health message about tanning and

sunscreen (no sugar message), and *arm 5* watched a placebo video about earthquakes (no sugar or health message).

Participants

We used the Prolific platform (Prolific Academic Ltd) [22] to recruit the study participants. Prolific is a web-based platform designed to connect researchers and individuals from different countries interested in participating in web-based academic research studies in exchange for payment. The main advantages of the platform are access to a diverse pool of web-based participants, affordability, and speed of recruitment [23]. Prolific implements several tools to reduce selection biases and allows researchers to specify various recruitment criteria, such as first language, age, sex, country of residence, and ethnicity, among others. Currently, the platform's participant pool consists of 150,000 individuals from 34 countries. Inclusion criteria in our study included being between the ages of 18 and 59 years (male, female, or other), being able to speak English, and having a residence in the United Kingdom. Exclusion criteria were not any of the inclusion criteria. The participants were not excluded based on an existing health condition (eg, diabetes) because Prolific does not collect health information from its users. Participants were provided with an informed consent form on the Prolific platform, which explained the purpose of the study, the risks and benefits of the research, and the means by which a participant could contact the researcher (and the human subjects review board at the Heidelberg University). After consenting, the Prolific platform redirected participants to the Gorilla platform (Cauldron Science Limited) [24], where the study was hosted. Gorilla is a cloud platform that provides versatile tools for web-based, experimental, and behavioral research [25]. The participants were also informed that they would be paid £1 (US $1.37) for the 10-minute completion time. We recruited participants until the target sample size was reached.

Procedures

Participants were asked basic demographic questions about their age, sex, and highest education level. The Gorilla algorithm then randomly assigned participants at a 1:1:1:1:1 ratio to the trial arms. The participants watched 1 video from start to finish.

The intervention video (*arms 1-3*) is a SAS video about sugar intake reduction developed by our coauthor (MA) at the Stanford School of Medicine [26-28]. It is animated, completely in English, and 3.42 minutes long. The video includes 2 main characters: a mother and her preadolescent daughter who are engaging in food-related activities, such as grocery shopping and cooking dinner. The video presents educational content on sugar-related health problems and includes a review of the World Health Organization recommendations for the daily consumption of added sugars. The narrative also mentions the girl's father who dies from diabetes complications because of frequent consumption of soda drinks.

The content placebo video is similar in style to the sugar intervention video—it is animated, has a length of 3.42 minutes, and a health message about the use of sunscreen and tanning [29]. We used the content placebo video to isolate the *content effect* of the sugar intervention video. As both the intervention and content placebo videos have a health message, we expect that any significant difference in reactance should be due to the sugar reduction content of the intervention video.

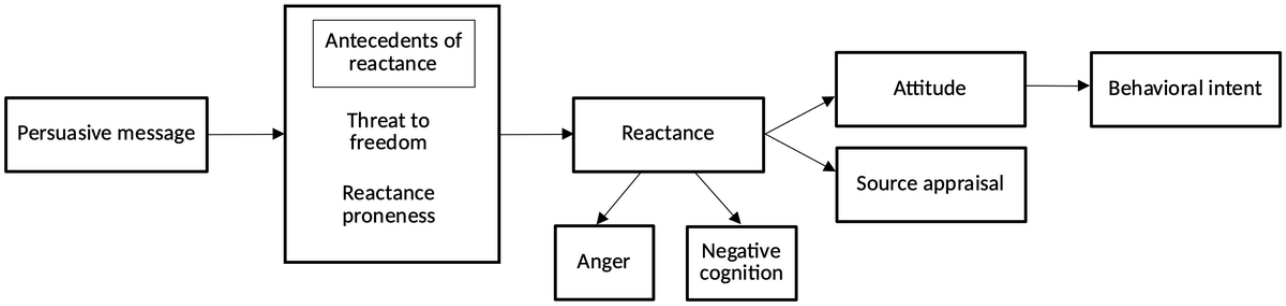
The placebo video [30] is also animated and has the same length as the intervention and content placebo videos. It describes the causes and characteristics of earthquakes, and contains no health-related or sugar consumption content. As the content placebo video promotes a health message and the placebo video does not, we expected the placebo video to arouse a very small (or even null) level of reactance. Thus, any significant difference in reactance levels between the content placebo and the placebo videos after randomization can be attributed to the content of the sunscreen message. We call this difference the *health awareness effect*. We describe the *total intervention effect* as the difference between the sugar intervention and the placebo videos, which is the sum of the content and health awareness effects.

The full explanation for the choice of comparators has been described in the study protocol [31].

Outcome Measures

The primary outcomes in this study were based on the Intertwined Process Cognitive-Affective Model from Dillard and Shen [5] and Zhang [32] (Figure 1). In this model, there are 2 antecedents to reactance (threat to freedom and trait proneness to reactance), reactance itself (consisting of anger and negative cognition), and its consequences (source appraisal, attitude, and behavioral intent). Reactance serves as a mediator between the antecedents of reactance and the behavioral intent to undertake the promoted health activity. In this paper, we focus on the antecedents of reactance (trait reactance proneness and threat to freedom), psychological reactance (consisting of anger and negative cognition), source appraisal, and attitude. All items were measured on a 5-point Likert scale (unless stated otherwise) with the following points: (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, and (5) strongly agree.

Figure 1. The Intertwined Process Cognitive-Affective Model, adapted from Dillard and Shen [5] and Zhang [32].



Trait Reactance Proneness

Trait reactance proneness refers to reactance being a personality attribute that causes the levels of experienced reactance to vary from individual to individual [33]. High-trait reactant individuals tend to experience reactance in certain situations and are more resistant to persuasion due to their strong need for independence and autonomy and a tendency to oppose authority [5,34].

Trait reactance proneness in this study was measured using the Hong Psychological Reactance Scale developed by Hong et al [33]. The scale consists of 11 items that comprise 4 major factors: emotional response to restricted choice, reactance to compliance, resisting influence from others, and reactance to advice and recommendations (Textbox 1).

Textbox 1. Trait reactance items based on the Hong Psychological Reactance Scale [33].

Emotional response to restricted choice

- 6. I become frustrated when I am unable to make free and independent decisions.
- 7. It irritates me when someone points out things which are obvious to me.
- 8. I become angry when my freedom of choice is restricted.

Reactance to compliance

- 1. Regulations trigger a sense of resistance in me.
- 2. I find contradicting others stimulating.
- 3. When something is prohibited, I usually think, “That’s exactly what I am going to do.”

Resisting influence from others

- 11. I resist the attempts of others to influence me.
- 12. It makes me angry when another person is held up as a role model for me to follow.
- 13. When someone forces me to do something, I feel like doing the opposite.

Reactance to advice and recommendations

- 5. I consider advice from others to be an intrusion.
- 9. Advice and recommendations usually induce me to do just the opposite.

Threat to Freedom

To measure the threat level of the message, we used the following 4 items from Dillard and Shen [5]:

1. The message threatened my freedom to choose.
2. The message tried to make a decision for me.
3. The message tried to manipulate me.
4. The message tried to pressure me.

Psychological Reactance

Following Dillard and Shen’s model, psychological reactance consists of 2 major components: (1) affective (anger) and (2) cognitive (negative cognition) [5]. Therefore, reactance was assessed by measuring the average of all items on the anger and negative cognition scales. To measure anger, the following 4 items were used:

1. This message makes me feel irritated.
2. This message makes me feel annoyed.
3. This message makes me feel aggravated.
4. This message makes me feel angry.

Negative cognition was measured using the scale from Quick et al [35] with the following 3 items:

1. The thoughts I had while watching this video were mostly unfavorable.
2. The thoughts I had while watching this video were mostly negative.

3. The thoughts I had while watching this video were mostly bad.

Source Appraisal

Source appraisal, also called source derogation [6], refers to the audience's evaluation of the source of the message. Source appraisal was examined using the question "The narrator of this video was..." and 7 semantic differential items anchored on either end with opposing adjectives: stupid or smart, unknowledgeable or knowledgeable, uninformed or informed, unintelligent or intelligent, unqualified or qualified, unreliable or reliable, and inexpert or expert [36]. The category ratings were scored from 1 to 5, with higher scores indicating more favorable evaluations of the message source (reverse-coded).

Attitude

Attitude toward message advocacy was measured using the following 4 items from Shen [37]:

1. I agree with what the message recommends.
2. I support what the message advocates.
3. I am in favor of the position in the message.
4. I endorse the claims made in the message.

Sample Size

We calculated the sample size needed for pairwise comparisons among the 5 groups using the analysis of variance. Our calculations resulted in a sample size of n=769 per group [31]. For a 5-way comparison, the sample size is N=3845. We selected a sample size of N=4000 to ensure we have sufficient power and account for attrition.

Statistical Methods

Descriptive statistics were used to obtain means and SDs of the demographic data of the sample, which included age, sex, and education status. We used analysis of variance to estimate the difference in the means of the outcome measures between the sugar intervention videos, the content placebo video, and the placebo video. The significance level α was set at .05. Post hoc tests with Tukey range method were used to create CIs for all pairwise differences between the means while controlling for family error rate. The placebo arm was chosen as the reference group, as the placebo video did not include any content related to sugar or health and, therefore, did not have any persuasive intent. All statistical analyses were performed using the statistical software R (R Foundation for Statistical Computing).

Availability of Data and Materials

The data were collected and stored on the Gorilla platform. The study investigators own and have complete control of the research data, which can be accessed at any time. For statistical analysis, the data were downloaded and stored safely in a computing system maintained by the Heidelberg University.

Ethical Approval

Ethical approval was obtained from the Heidelberg University's ethics committee (Universität Heidelberg Ethikkommission der Medizinische Fakultät) on March 18, 2020, protocol: S-088/2020.

Results

Sample Characteristics

Between December 9, 2020, and December 11, 2020, a total of 4159 participants from the United Kingdom were recruited for the trial. After recruitment, 0.26% (11/4159) participants were lost and another 3.24% (135/4159) participants were dropped, as they did not complete the study for either technical reasons (poor internet connection, video loading issues, system crash, and so on) or other unknown reasons. Of the recruited sample, 96.48% (4013/4159) completed the trial and were included in the final analysis (Figure 2). Table 1 provides the demographic characteristics of the participants by group, including gender, age, and education level. Of the sample, 60.90% (2444/4013) were female, 32.27% (1295/4013) were aged between 25 and 34 years, and 64.09% (2572/4013) had some college education or a bachelor's degree. There were no significant differences in baseline characteristics between the 5 arms, suggesting that the randomization was efficient.

Figure 2. Trial design.

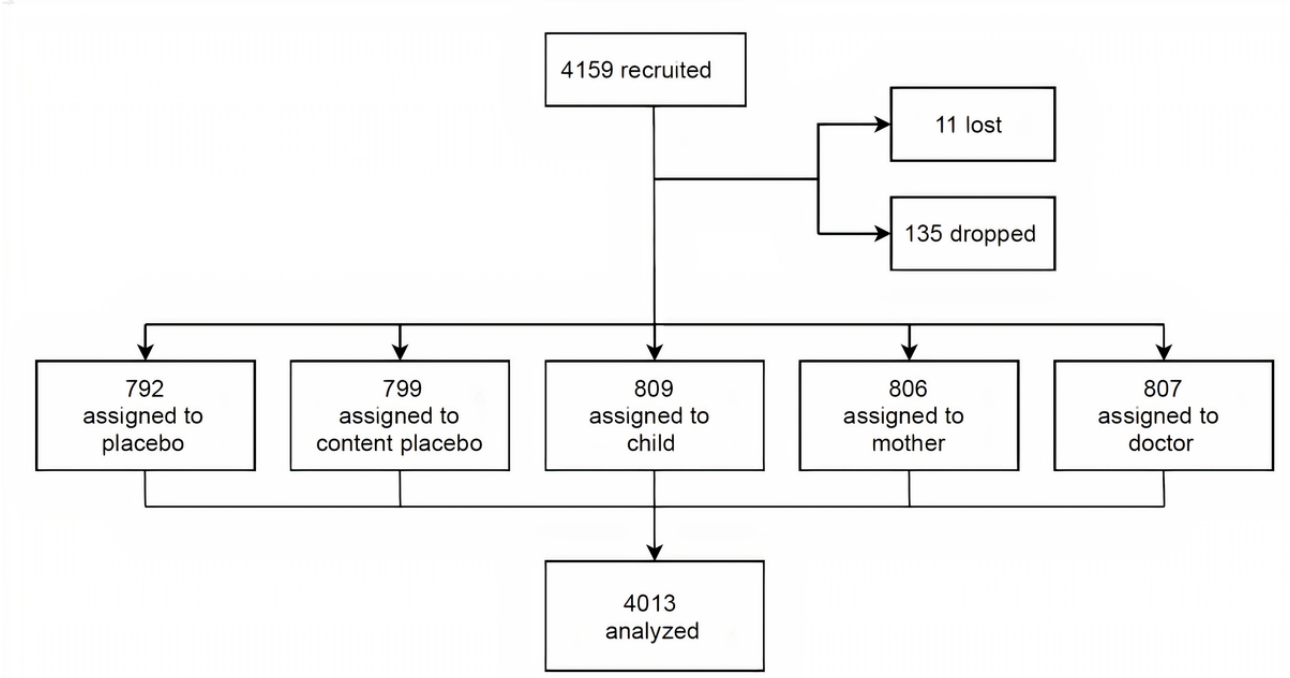


Table 1. Summary of demographic characteristics by group (N=4013).

Characteristics	Placebo (n=792), n (%)	Content placebo (n=799), n (%)	Child voice (n=809), n (%)	Mother voice (n=806), n (%)	Physician voice (n=807), n (%)	*P* value
Gender						.83
Female	485 (61.24)	481 (60.20)	494 (61.06)	485 (60.17)	497 (61.59)	
Male	300 (37.88)	313 (39.17)	313 (38.69)	317 (39.33)	307 (38.04)	
Other	7 (0.88)	5 (0.63)	2 (0.25)	4 (0.50)	3 (0.37)	
Age (years)						.96
18-24	208 (26.26)	184 (23.03)	214 (26.45)	200 (24.81)	195 (24.16)	
25-34	250 (31.57)	259 (32.42)	266 (32.88)	267 (33.13)	254 (31.47)	
35-44	167 (21.09)	175 (21.90)	175 (21.63)	167 (20.72)	190 (23.54)	
45-54	120 (15.15)	130 (16.27)	109 (13.47)	121 (15.01)	120 (14.87)	
55-59	47 (5.93)	51 (6.38)	45 (5.56)	51 (6.33)	48 (5.95)	
Education						.97
Primary School or less	11 (1.39)	13 (1.63)	8 (0.99)	9 (1.12)	10 (1.24)	
Completed High School	126 (15.91)	123 (15.39)	117 (14.46)	131 (16.25)	126 (15.61)	
Some College, Bachelor's Degree	500 (63.13)	501 (62.70)	530 (65.51)	518 (64.27)	525 (65.06)	
Master's Degree, Doctorate	155 (19.57)	162 (20.27)	154 (19.04)	148 (18.36)	146 (18.09)	

Outcome Measures

Table 2 presents the descriptive statistics, including the mean and SDs of all the key variables measured in this study.

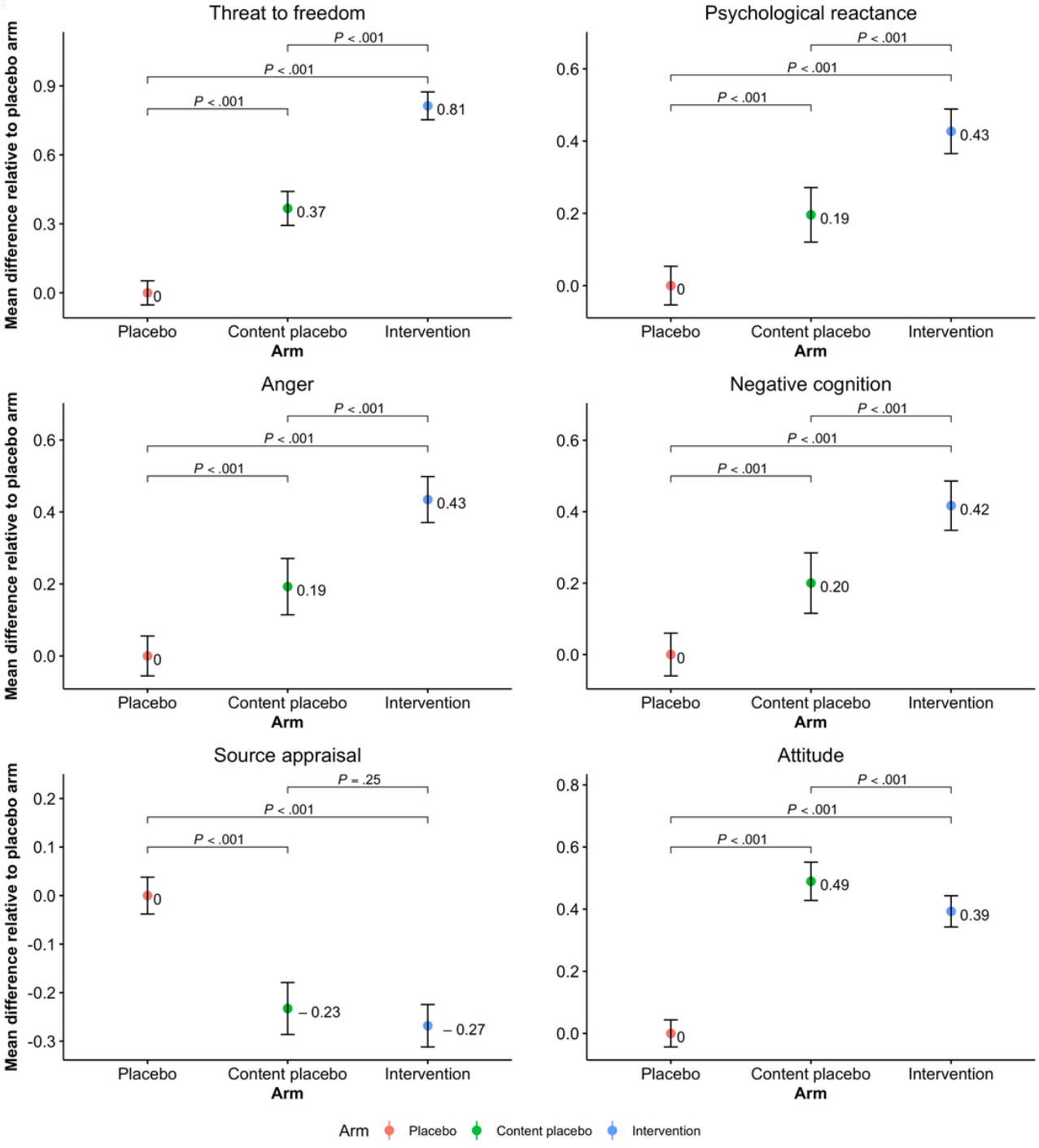
Table 2. Mean and SD of outcome variables in study arms.

Characteristics	Placebo (n=792), mean (SD)	Content placebo (n=799), mean (SD)	Child voice (n=809), mean (SD)	Mother voice (n=806), mean (SD)	Physician voice (n=807), mean (SD)	*P* value
Trait reactance proneness	2.98 (0.48)	2.97 (0.51)	2.99 (0.52)	2.97 (0.54)	2.97 (0.52)	.92
Threat to freedom	1.46 (0.55)	1.83 (0.68)	2.28 (0.86)	2.34 (0.82)	2.20 (0.80)	<.001
Psychological reactance	1.56 (0.63)	1.76 (0.69)	1.99 (0.87)	2.03 (0.83)	1.95 (0.79)	<.001
Anger	1.51 (0.63)	1.70 (0.72)	1.95 (0.90)	1.98 (0.87)	1.90 (0.83)	<.001
Negative cognition	1.64 (0.77)	1.84 (0.79)	2.05 (0.95)	2.09 (0.91)	2.02 (0.86)	<.001
Attitude	3.79 (0.60)	4.28 (0.60)	4.22 (0.64)	4.14 (0.64)	4.18 (0.65)	<.001
Source appraisal	3.91 (0.52)	3.67 (0.52)	3.57 (0.58)	3.63 (0.56)	3.72 (0.54)	<.001
Stupid or smart	3.77 (0.78)	3.50 (0.79)	3.74 (0.81)	3.59 (0.80)	3.62 (0.75)	<.001
Unknowledgeable or knowledgeable	4.06 (0.71)	3.93 (0.75)	3.82 (0.81)	3.86 (0.74)	3.88 (0.67)	<.001
Uninformed or informed	4.17 (0.59)	4.08 (0.66)	4.03 (0.69)	4.02 (0.68)	4.03 (0.62)	<.001
Unintelligent or intelligent	3.96 (0.63)	3.69 (0.67)	3.80 (0.67)	3.72 (0.66)	3.72 (0.65)	<.001
Unqualified or qualified	3.75 (0.71)	3.44 (0.68)	3.06 (0.94)	3.34 (0.74)	3.56 (0.72)	<.001
Unreliable or reliable	3.91 (0.67)	3.69 (0.62)	3.57 (0.75)	3.64 (0.72)	3.75 (0.66)	<.001
Inexpert or expert	3.71 (0.71)	3.40 (0.65)	2.99 (0.86)	3.23 (0.73)	3.45 (0.69)	<.001

Antecedents of Reactance

Trait reactance proneness and threat to freedom are antecedents to reactance. Higher scores on trait proneness and threat to freedom scales indicate higher proneness to reactance and greater perceived threat, respectively. As shown in Table 2, the mean scores for trait reactance proneness in the 5 arms were in the mean range 2.97-2.99 (SD 0.48-0.52) with *P*=.92, which did not vary significantly between the 5 arms. When comparing the means scores for threat to freedom, the analysis revealed a significant difference between the 5 groups (*P*<.001). Furthermore, the pairwise comparisons for the threat level, using a Bonferroni correction, indicated that participants in the content placebo and intervention arms reported higher threat to freedom when compared with the placebo arm (*P*<.001; Figure 3). When comparing intervention arms between each other, participants in the mother arm indicated a higher threat level than those in the doctor arm (*P*=.002) but not in the child arm (*P*=.52). Although the threat level in the child arm (mean 2.28, SD 0.86) was slightly higher than in the doctor arm (mean 2.20, SD 0.80), the difference was not found to be significant (*P*=.21).

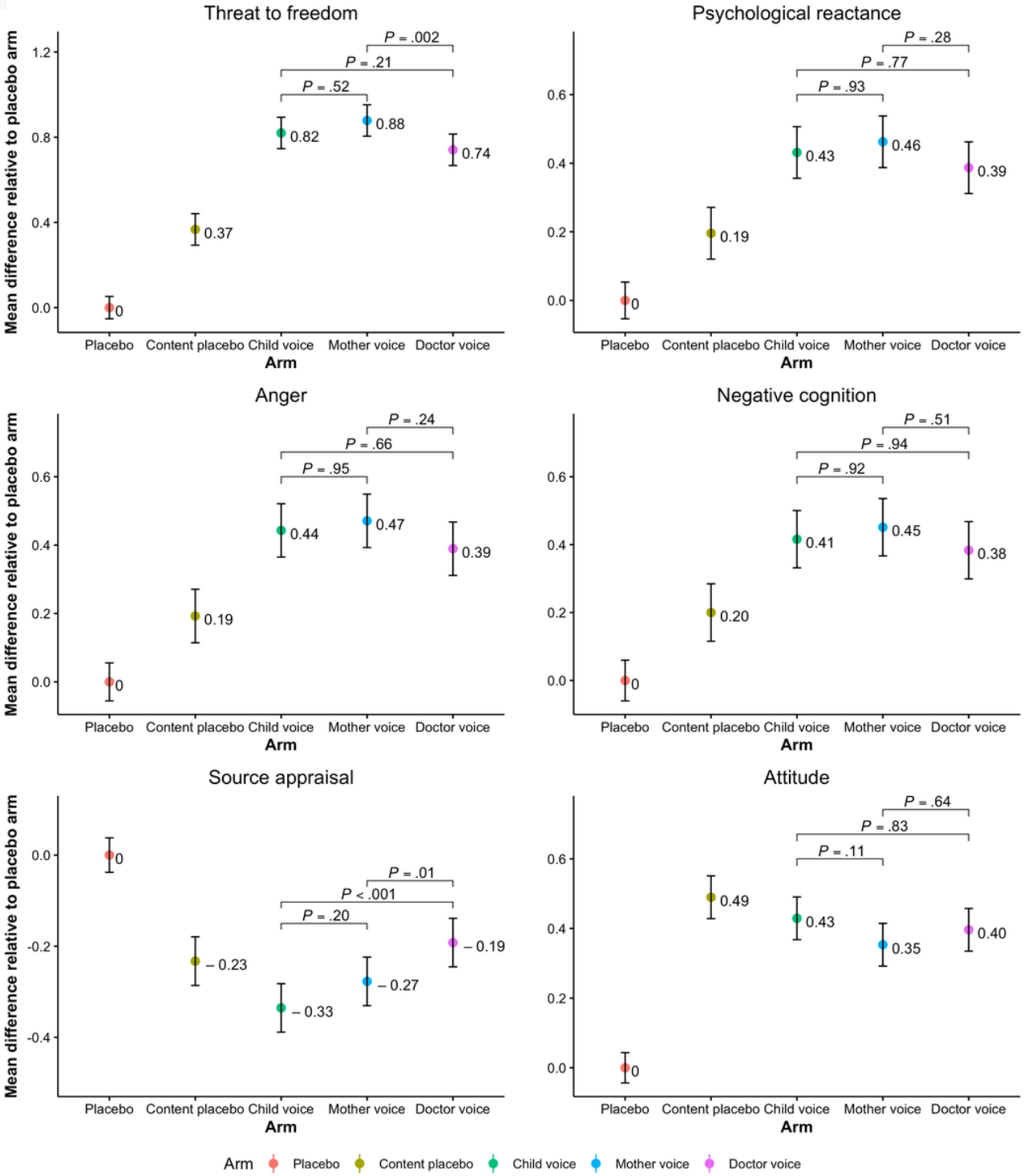
Figure 3. Mean differences in outcome measures among the placebo arm (reference arm), content placebo, and intervention arms. *P* values represent the significance of the observed difference in means among the study arms.



Psychological Reactance

Psychological reactance was assessed by measuring anger and negative cognition. Therefore, the average of all items on anger and negative cognition indicated the total score on reactance. Higher scores on anger implied a greater level of anger, whereas higher scores on negative cognition suggested a higher presence of negative thoughts following the video. Therefore, it was expected that higher scores on reactance would indicate higher levels of reactance triggered by the video. A 5-group comparison revealed a significant difference in the reactance levels (P<.001). Figure 3 shows that when compared with the placebo arm, the content placebo and intervention arms had significantly higher scores on reactance (P<.001). However, there was no statistically significant difference between intervention arms, suggesting that participants experienced the same amount of reactance when watching the video narrated by either the child, mother, or physician. When considering anger and negative cognition scales separately, the analysis revealed similar outcomes, where all arms were significantly different when compared with the placebo arm (Figure 3) and the intervention arms did not differ from each other (Figure 4).

Figure 4. Mean differences in outcome measures among the placebo arm (reference arm), content placebo, and the 3 intervention arms (child, mother, and physician). *P* values represent the significance of the observed difference in means between the intervention arms.



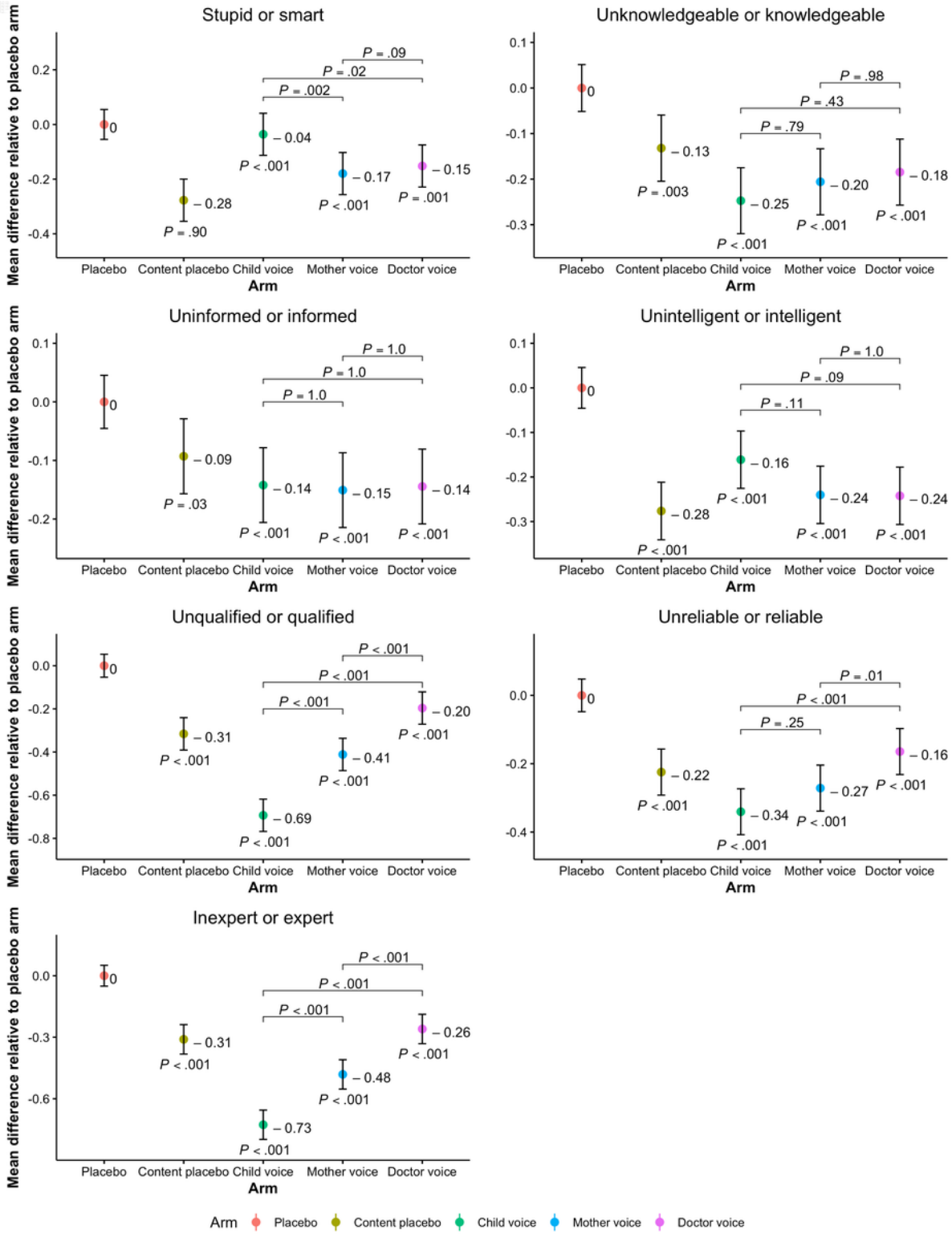
Source Appraisal

The evaluation of the message source was considered more favorable when the participants scored high on the source appraisal scale. The difference in means was significant between all the study arms (*P*<.001). As seen in Figure 3, the pairwise comparisons generated significant differences in mean scores between placebo and content placebo (*P*<.001) as well as placebo and intervention arms (<.001), suggesting that participants in the placebo arm had a more favorable evaluation of the video narrator than the participants in other study arms. However, there was no significant difference in the appraisal of the source between the content placebo and intervention arms (P=.25). When comparing intervention arms separately, participants who watched the video narrated by the physician had higher scores, that is, a more positive evaluation of the source, than the participants who watched videos narrated by the child (*P*<.001) or the mother (P=.01; Figure 4).

Items on the source appraisal scale were also analyzed separately for a more detailed understanding of the message source

evaluation as source (narrator) is the major component of our research question. As seen in Figure 5, the child narrator was considered to be smarter than mother (P=.002) and physician (P=.03). In addition, unlike the narrators in the content placebo (P<.001), mother (P<.001), and doctor (P<.001) arms, the narrator in the child arm was found to be as smart as the narrator in the placebo arm (P=.89). However, all 3 intervention arms (P<.001) and the content placebo arm (P=.003) scored significantly lower on the unknowledgeable or knowledgeable item in comparison with the placebo arm. There were no significant differences between intervention arms on this item. The placebo arm narrator was also seen to be more informed than the narrators in the content placebo (P=.04) and intervention arms (P<.001; Figure 5). The narrator in the placebo arm was also believed to be more intelligent, qualified, reliable, and expert than the narrators in the other 4 arms (P<.001). When comparing the intervention arms with each other, all narrators were considered to be equally informed and intelligent. However, the participants in the doctor arm rated the message source to be more qualified (P<.001), reliable (P<.001 and P=.02), and expert (P<.001) than those in the child and mother arms, respectively. Furthermore, the mother narrator was also seen as more qualified (P<.001) and expert (P<.001) than the child narrator, although the same difference was not found for the unreliable or reliable item (P=.25; Figure 5).

Figure 5. Source appraisal scale items. The y-axis shows the mean difference with CIs of the items in the content placebo, child, mother, and doctor arms relative to the placebo arm (reference arm). The x-axis shows the trial arms. *P* values under CIs represent the significance of the observed difference in means relative to the placebo arm, and *P* values over brackets represent the significance of the observed difference in means among the intervention arms.



Attitude

Higher scores on the attitude scale suggested that participants had a more favorable attitude toward the message advocacy. After the analysis showed a significant difference on the attitude score between the study arms (*P*<.001), pairwise comparisons confirmed that participants in the content placebo and intervention arms had significantly more positive attitudes toward the message when compared with participants in the placebo arm (*P*<.001; Figure 3). Furthermore, compared with the content placebo, participants in the intervention arm had significantly less favorable attitude toward the sugar videos (*P*<.001). As the message in all 3 intervention videos was the

same, there were no statistically significant differences between the child, mother, and doctor arms (Figure 4).

Discussion

Principal Findings

In recent years, the use of a narrative-structured format and video-based animation has enabled the creative use of nonhuman and nonadult characters to promote persuasive health messages [38]. To further explore the design of effective health communication interventions and increase their long-term effects, we created a SAS video that can engage global audiences in evidence-based health promotion and be rapidly distributed on various social media platforms. In this study, we evaluated whether a SAS video about sugar intake could attenuate reactance to a health message (hypothesis 1) and compared the effectiveness of a child narrator, her mother, and physician in reducing reactance to a message about added sugar consumption (hypothesis 2).

We found that our SAS video aroused higher levels of reactance when compared with a content placebo video about sunscreen use (containing no sugar message) and a placebo video about earthquakes (containing no health message). With respect to our first hypothesis, we therefore demonstrate a causal relationship between exposure to a SAS video and the antecedents and components of reactance. In particular, our results show that compared with the placebo video, the content placebo video was perceived as more threatening, while the intervention videos were seen as the most threatening. In addition, participants who watched the intervention videos experienced significantly higher levels of anger and negative cognition than those in the placebo and content placebo arms. Although psychological reactance has not been fully tested in the context of digital health promotion, this study contributes evidence to the existing literature on persuasion and reactance [39].

One plausible explanation for the higher levels of reactance to the intervention videos may be rooted in the part that describes the death of the child's father, which is attributed to the regular consumption of soda drinks. Some scholars have argued that health promotion messages can be effective when they increase people's fear or concern about risky behaviors that can threaten their health [40,41]. In other words, when presented with fear, individuals may be more motivated to change their behaviors. The cognitive functional model by Nabi [42] suggests that before creating a health message, authors should determine which emotion would be most suited to their persuasive goals, adding that fear might be best used for preventing behaviors that lead to severe consequences, such as death. However, participants may have perceived this part of the story as an emotional manipulation and recognized the actual persuasive intention in the message. Some scholars have concluded that noticing a covert attempt to promote healthy behavior disguised as entertainment results in reactance, whereas a more direct persuasive attempt does not [7]. Several other studies [43,44] have also shown that health advertising material that was perceived to be manipulative caused more resistance and anger and was, therefore, less effective in changing attitudes. Taking these points into account, the removal of the emotional part of the video, where the death of the child's father is portrayed, could potentially induce less reactance and negative attitudes toward the message. This presents a possible avenue for future research, in which we could compare videos with and without this emotional subplot.

In this study, we focused on 1 modifiable component—the social authority of the narrator—and its effect on reactance to a message about reducing sugar intake. Initially, we assumed that a child narrator would be a more persuasive messenger and would be less likely to arouse reactance when compared with the adult narrators (mother and physician), as the audience may view the child as nonthreatening and lacking vested interest. Contrary to our expectations, we found no evidence that the child narrator attenuated reactance when compared with the same intervention videos narrated by the mother and the family physician. Therefore, we fail to reject the null hypothesis of H2 in that there were no significant differences between the child and the mother and the child and the physician with respect to the threat to freedom, anger, negative cognition, and the combination of the last 2 components (state reactance).

However, one of the few significant differences that we observed was in the source appraisal component, which plays a critical role in this context. Our results show that the physician narrator was perceived to be more qualified and reliable and has more expertise than the child and the mother narrators. This is an expected finding as doctors are generally seen as experts in the health field and reliable sources of accurate and valid information. Earlier studies have shown that message recipients tend to be more motivated to change and persuaded by an expert rather than a non–expert message source [45,46]. However, it has also been suggested that the position of the person toward the message, that is, whether it is viewed as consistent with or discrepant from one's current attitude toward the issue, as well as issue relevancy determine the persuasive effect of the message, regardless of the source expertise [47]. One study [48] found that when participants had low relevance to the issue, higher source expertise produced better attitudes toward the argument, even when the message quality was manipulated. However, when the message had high relevance, message quality had the biggest impact on the attitudes of the participants, whereas source expertise became a less important factor of persuasion. Our study provides further evidence to these findings as the videos narrated by child, mother, and physician produced similar reactance outcomes, suggesting that the message itself and its relevance may play a bigger role than the source of the message.

Another minor yet interesting finding was that the participants evaluated the child source as smarter than the mother and the physician. A possible explanation for this difference might lie in the fact that participants, who were exclusively adults, were not willing to call the child narrator stupid owing to potential social desirability bias, which is described as the tendency of research participants to give socially desirable responses instead of honest responses [49]. As we could not find any prior studies that compared child narrators to adult narrators, there is no evidence to support this assumption.

We do not believe that design differences between the intervention and placebo videos can account for differences in reactance across the trial arms. This is because we were careful to select placebo videos that were similar to the sugar intervention video, such that all videos were short (3.42 minutes), animated, story-based, and in English. The only systematic difference among the videos was the content of the narrated messages (about earthquakes, sunscreen, and sugar), which were the trial arms. Importantly, as we explain in Figure 1, the antecedents to reactance are *threat to freedom* and *trait proneness to reactance*. This means that design differences such as animation style, background shapes or colors, and target audience are not hypothesized to arouse reactance and hence are unlikely to account for differences in reactance. Nevertheless, we acknowledge that the sugar intervention video was narrated by a female voice, and the placebo videos were narrated by male voices. There is evidence that men are perceived as more credible than women, and women are perceived as more trustworthy than men [50]. To the best of our knowledge, there is no evidence that links the gender of the narrator to differences in reactance, which could be a future avenue of research.

The findings of this study make an important contribution to the literature on digital health promotion. Several studies that focused on added sugar reduction have used nonanimated, web-based videos, such as a puppet show [51], expert opinion intercut with case studies [52], video courses [53], and storytelling interviews [54]. Unlike the SAS video in our study, these interventions were approximately 6 to 15 minutes long, which is longer than the optimal time required for a social media format and focused on certain demographic groups and populations. Although we could not support the proposed hypothesis that a child can be a powerful and persuasive health promotion agent, our findings indicate that the quality and design of the health message should be considered more carefully in persuasive health promotion. The finding that a message with a persuasive intent, even when masked, provokes some kind of reactance may be reasonable, but the end goal of health promotion experts should be to create and promote SAS videos that would lead to a minimum amount of reactance and be almost comparable with a message in which persuasion is absent. The avoidance of intense emotional appeal and the use of narrative-based messages could be potentially successful components in health message design.

Strengths and Limitations

This study had several strengths. First, we used an RCT design, which allowed us to eliminate any potential sources of bias by randomization. Randomization ensures that there are no systematic differences introduced at the enrollment stage. Second, the web-based nature of our experiment enabled us to reach a large sample size, which ensured the quality and reliability of the sample. The use of content placebo and placebo videos is also an innovative feature of our study, which enabled us to isolate the health awareness effect and content effect of the intervention video. We are not aware of any previous study that had such a large sample size and used a similar experimental approach to examine the social authority of the health message source. Once the design of the SAS video that we created is further examined and modified, it can be used for larger audiences on social media and other educational sources, as it is short, simple, and quickly scalable.

Our study has several limitations. First, Prolific uses convenience sampling to recruit participants, so that study places are filled on a first-come, first-serve basis. Thus, a considerable portion of responses could come from participants who are on the web at the time a study is launched or immediately afterward. Rapid-responder bias may be an issue if the required sample is very small or very specific. However, our study was general (men, women, other; aged 18-59 years; of any education level if a UK resident) and ran continuously for 3 consecutive days. In addition, Prolific has several mechanisms to reduce rapid-responder bias and equally distribute study places among active participants. For example, when a study is launched, Prolific sends an email to a random subset of all eligible participants every 48 hours until the sample size is reached. Therefore, it is unlikely that rapid-responder bias will have significantly affected our results. Second, our study had a sampling bias toward women (60.85%, 2442/4013 females vs 38.62%, 1550/4013 males) and participants with higher education (83.20%, 3339/4013) had a bachelor's degree or higher). Similar sample distributions have been reported in several web-based studies [55,56]. It has been observed that most participant pools in the social sciences are biased toward Western, educated, industrialized, rich, and democratic individuals, as they are predominantly from the United Kingdom, United States, or Europe [57]. The generalizability of our findings may therefore be limited to the UK and the US contexts, and possibly to Europe. Further research should be conducted in other settings to make the results more generalizable to other geographies and cultures.

A third limitation is that the Prolific participants may have chosen to participate in our study because of the sugar-related topic. However, it is unlikely that this type of selection bias would have considerably affected our results for one important reason: randomization. For our RCT, we randomized participants to either the sugar intervention video, the content placebo video, or the placebo video, so that any topic-specific selection bias would have been uniformly distributed across the trial arms. Fourth, it is possible that participants may have been motivated by financial rewards, which could introduce a selection bias. Again, this form of bias would have been equally distributed across the trial arms because of our randomized design. In addition, financial rewards are standard for web-based studies and are acceptable when the research is not focused on a specific disease or treatment and does not involve potential risks [58]. The study reward was also relatively small (£1; US $1.37) and most of the participants were highly educated (most had a bachelor's degree or higher), making it unlikely that participation was motivated by economic disadvantage. Indeed, previous research has reported that web-based research participants are motivated by a variety of reasons other than financial rewards, such as self-improvement, microtasking to avoid wasted time, and other emotional benefits [58]. Overall, it is unlikely that a small financial reward led to biases that significantly affected our results. Finally, although this was beyond the scope of our study, we acknowledge that further

research is needed to determine if SAS video interventions for social media can be more cost-effective than mobile health or other nondigital approaches [59].

Taken together, the findings of this study demonstrate that the content of health messages may have a greater impact on reactance than the source of the message and its authority. Moreover, the use of SAS videos on social media can facilitate public health efforts to promote healthy behaviors to a larger audience. The experimental design and the use of multiple placebo groups in this study present novel approaches for further investigation in this area. It is essential to gain a better understanding of the unique ways in which messages about healthy behaviors are processed and the resulting emotions and intentions to advance health communication strategies worldwide.

Acknowledgments

This study was funded by an Alexander von Humboldt University Professor Prize awarded to TB.

Authors' Contributions

VH and AV wrote the manuscript. MA designed, produced, and created all 3 sugar videos (child, mother, and physician). AV, VH, and CF contributed to the questionnaire development. AV and TB designed the trial. All authors provided comments and feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1011 KB - jmir_v23i11e29664_app1.pdf]

References

1. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):186-194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
2. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
3. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
4. Singhal A, Rogers E. Entertainment-Education: A Communication Strategy for Social Change. Mahwah, NJ, USA: Erlbaum Associates; 1999:1-280.
5. Dillard JP, Shen L. On the nature of reactance and its role in persuasive health communication. Commun Monogr 2005 Jun;72(2):144-168. [doi: 10.1080/03637750500111815]
6. Miller CH, Lane LT, Deatrick LM, Young AM, Potts KA. Psychological reactance and promotional health messages: the effects of controlling language, lexical concreteness, and the restoration of freedom. Human Comm Res 2007 Apr;33(2):219-240 [FREE Full text] [doi: 10.1111/j.1468-2958.2007.00297.x]
7. Moyer-Gusé E, Nabi R. Explaining the effects of narrative in an entertainment television program: overcoming resistance to persuasion. Hum Commun Res 2010;36(1):26-52. [doi: 10.1111/j.1468-2958.2009.01367.x]
8. Vandormael A, Adam M, Greuel M, Gates J, Favaretti C, Hachaturyan V, et al. The effect of a wordless, animated, social media video intervention on COVID-19 prevention: online randomized controlled trial. JMIR Public Health Surveill 2021 Jul 27;7(7):e29060 [FREE Full text] [doi: 10.2196/29060] [Medline: 34174778]
9. Brehm SS, Brehm JW. Psychological Reactance: A Theory of Freedom and Control. Cambridge, MA, United States: Academic Press; 1981:432.
10. Brewer NT, Jeong M, Hall MG, Baig SA, Mendel JR, Lazard AJ, et al. Impact of e-cigarette health warnings on motivation to vape and smoke. Tob Control 2019 Jul 10;28(e1):64-70 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054878] [Medline: 31292169]
11. Cingolani AM, Barberá I, Renison D, Barri FR. Can persuasive and demonstrative messages to visitors reduce littering in river beaches? Waste Manag 2016 Dec;58:34-40. [doi: 10.1016/j.wasman.2016.08.028] [Medline: 27613415]
12. Quick BL, Bates BR. The use of gain- or loss-frame messages and efficacy appeals to dissuade excessive alcohol consumption among college students: a test of psychological reactance theory. J Health Commun 2010 Sep 30;15(6):603-628. [doi: 10.1080/10810730.2010.499593] [Medline: 20812123]

13. Bullock K, Lahne J, Pope L. Investigating the role of health halos and reactance in ice cream choice. Food Qual Prefer 2020 Mar;80:103826. [doi: 10.1016/j.foodqual.2019.103826]
14. Chen M, McGlone MS, Bell RA. Persuasive effects of linguistic agency assignments and point of view in narrative health messages about colon cancer. J Health Commun 2015 Aug 18;20(8):977-988. [doi: 10.1080/10810730.2015.1018625] [Medline: 26087078]
15. Elbert SP, Dijkstra A, Rozema AD. Effects of tailoring ingredients in auditory persuasive health messages on fruit and vegetable intake. Psychol Health 2017 Jul 07;32(7):781-797. [doi: 10.1080/08870446.2017.1300259] [Medline: 28266878]
16. Gardner L, Leshner G. The role of narrative and other-referencing in attenuating psychological reactance to diabetes self-care messages. Health Commun 2016 Nov 03;31(6):738-751. [doi: 10.1080/10410236.2014.993498] [Medline: 26528578]
17. Avery EJ. The role of source and the factors audiences rely on in evaluating credibility of health information. Public Relat Rev 2010 Mar;36(1):81-83. [doi: 10.1016/j.pubrev.2009.10.015]
18. Mogre V, Scherpbier AJ, Stevens F, Aryee P, Cherry MG, Dornan T. Realist synthesis of educational interventions to improve nutrition care competencies and delivery by doctors and other healthcare professionals. BMJ Open 2016 Oct 21;6(10):e010084 [FREE Full text] [doi: 10.1136/bmjopen-2015-010084] [Medline: 27797977]
19. Kareklas I, Muehling DD, Weber T. Reexamining health messages in the digital age: a fresh look at source credibility effects. J Advert 2015 May 05;44(2):88-104. [doi: 10.1080/00913367.2015.1018461]
20. Vollbrecht JL, Roloff ME, Paulson GD. Coercive potential and face-sensitivity: the effects of authority and directives in social confrontation. Int Jnl of Conflict Management 1997 Mar;8(3):235-251. [doi: 10.1108/eb022797]
21. Jung T, Shim W, Mantaro T. Psychological reactance and effects of social norms messages among binge drinking college students. J Alcohol Drug Educ 2010;54(3):7-18 [FREE Full text]
22. Prolific. URL: https://prolific.co/ [accessed 2021-08-18]
23. Lumsden J, Damer E. So, you want to recruit participants online? Prolific. 2018. URL: https://blog.prolific.co/so-you-want-to-recruit-participants-online/ [accessed 2021-08-25]
24. Gorilla. URL: https://app.gorilla.sc/ [accessed 2021-08-18]
25. Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK. Gorilla in our midst: an online behavioral experiment builder. Behav Res Methods 2020 Feb 23;52(1):388-407 [FREE Full text] [doi: 10.3758/s13428-019-01237-x] [Medline: 31016684]
26. Adam M. SRV_C_F. URL: https://www.youtube.com/watch?v=VaQsJwFOzmM [accessed 2021-08-18]
27. Adam M. SRV_M_F. URL: https://www.youtube.com/watch?v=SU-7tRWP238 [accessed 2021-08-18]
28. Adam M. SRV_P_F. URL: https://www.youtube.com/watch?v=PFCkT1fWbT8 [accessed 2021-08-18]
29. Adam M. Video B. URL: https://www.youtube.com/watch?v=CpS_E_znQzU [accessed 2021-08-19]
30. What is an earthquake? The Dr. Binocs show - Educational videos for kids. Peekaboo Kidz. URL: https://www.youtube.com/watch?v=dJpIU1rSOFY [accessed 2021-08-19]
31. Vandormael A, Adam M, Hachaturyan V. Reactance to social authority in entertainment-education media: protocol for a web-based randomized controlled trial. JMIR Res Protoc 2021;10(5):e25343. [doi: 10.2196/preprints.25343]
32. Zhang X. Effects of freedom restoration, language variety, and issue type on psychological reactance. Health Commun 2020 Oct 21;35(11):1316-1327. [doi: 10.1080/10410236.2019.1631565] [Medline: 31225759]
33. Hong S, Faedda S. Refinement of the Hong Psychological Reactance Scale. Educ Psychol Meas 2016 Jul 02;56(1):173-182. [doi: 10.1177/0013164496056001014]
34. Seibel CA, Dowd ET. Personality characteristics associated with psychological reactance. J Clin Psychol 2001 Jul;57(7):963-969. [doi: 10.1002/jclp.1062] [Medline: 11406807]
35. Quick BL, Kam JA, Morgan SE, Liberona CA, Smith RA. rospect theory, discrete emotions, and freedom threats: an extension of psychological reactance theory. J Commun 2014 Dec 13;65(1):40-61. [doi: 10.1111/jcom.12134]
36. Jenkins M, Dragojevic M. Pexplaining the process of resistance to persuasion. Commun Res 2011 Oct 17;40(4):559-590. [doi: 10.1177/0093650211420136]
37. Shen L. Antecedents to psychological reactance: the impact of threat, message frame, and choice. Health Commun 2015 Sep 25;30(10):975-985. [doi: 10.1080/10410236.2014.910882] [Medline: 25256920]
38. Adam M, Tomlinson M, Le Roux I, LeFevre AE, McMahon SA, Johnston J, et al. The Philani MOVIE study: a cluster-randomized controlled trial of a mobile video entertainment-education intervention to promote exclusive breastfeeding in South Africa. BMC Health Serv Res 2019 Apr 02;19(1):211 [FREE Full text] [doi: 10.1186/s12913-019-4000-x] [Medline: 30940132]
39. Braddock K, Dillard JP. Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. Commun Monogr 2016 Feb 25;83(4):446-467. [doi: 10.1080/03637751.2015.1128555]
40. Cameron L, Chan C. Designing health communications: harnessing the power of affect, imagery, and self-regulationlf-regulation and health communications. Soc Personal Psychol Compass 2008;2(1):262-282. [doi: 10.1111/j.1751-9004.2007.00057.x]
41. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]

42. Nabi RL. A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. Commun Theory 1999 Aug;9(3):292-320. [doi: 10.1111/j.1468-2885.1999.tb00172.x]
43. Fukada H. Psychological processes mediating persuasion-inhibiting effect of forewarning in fear-arousing communication. Psychol Rep 1986 Feb 31;58(1):87-90. [doi: 10.2466/pr0.1986.58.1.87] [Medline: 3961085]
44. Coulter RH, Pinto MB. Guilt appeals in advertising: what are their effects? J Appl Psychol 1995 Dec;80(6):697-705. [doi: 10.1037/0021-9010.80.6.697] [Medline: 8557622]
45. Hovland C, Weiss W. The influence of source credibility on communication effectiveness. Public Opin Q 1951;15(4):635-650. [doi: 10.1086/266350]
46. Jones L, Sinclair R, Courneya K. The effects of source credibility and message framing on exercise intentions, behaviors and attitudes: an integration of the elaboration likelihood model and prospect theory. J Appl Soc Psychol 2003;33(1):179-196. [doi: 10.1111/j.1559-1816.2003.tb02078.x]
47. Clark JK, Wegener DT, Habashi MM, Evans AT. Source expertise and persuasion: the effects of perceived opposition or support on message scrutiny. Pers Soc Psychol Bull 2012 Jan 30;38(1):90-100. [doi: 10.1177/0146167211420733] [Medline: 21878613]
48. Petty RE, Cacioppo JT, Goldman R. Personal involvement as a determinant of argument-based persuasion. J Pers Soc Psychol 1981;41(5):847-855. [doi: 10.1037//0022-3514.41.5.847]
49. Grimm P. Social desirability bias. In: Sheth J, Malhotra N, editors. Wiley International Encyclopedia of Marketing. Hoboken, New Jersey, United States: John Wiley & Sons, Ltd; 2010.
50. Weibel D, Wissmath B, Groner R. How gender and age affect newscasters' credibility—an investigation in Switzerland. J Broadcast Electron Media 2008 Aug 08;52(3):466-484. [doi: 10.1080/08838150802205801]
51. Glorioso I, Narciso Z, Avilla J, Capanzana M. Evaluation of a puppet video on physical activity as a diabetic education material for older children. Philipp J Sci 2018;147(4):589-596 [FREE Full text]
52. Dyson P, Beatty S, Matthews D. An assessment of lifestyle video education for people newly diagnosed with type 2 diabetes: video education for people with newly diagnosed type 2 diabetes. J Hum Nutr Diet 2010;23(4):353-359. [doi: 10.1111/j.1365-277x.2010.01077.x]
53. Chen Y, Huang Y, Chien Y, Chen MC. The effect of an online sugar fact intervention: change of mothers with young children. Nutrients 2020 Jun 22;12(6):1859 [FREE Full text] [doi: 10.3390/nu12061859] [Medline: 32580451]
54. Njeru JW, Patten CA, Hanza MM, Brockman TA, Ridgeway JL, Weis JA, et al. Stories for change: development of a diabetes digital storytelling intervention for refugees and immigrants to minnesota using qualitative methods. BMC Public Health 2015 Dec 29;15(1):1311 [FREE Full text] [doi: 10.1186/s12889-015-2628-y] [Medline: 26715465]
55. Jang M, Vorderstrasse A. Socioeconomic status and racial or ethnic differences in participation: web-based survey. JMIR Res Protoc 2019 Apr 10;8(4):e11865 [FREE Full text] [doi: 10.2196/11865] [Medline: 30969173]
56. Geldsetzer P. Knowledge and perceptions of COVID-19 among the general public in the United States and the United Kingdom: a cross-sectional online survey. Ann Intern Med 2020 Mar 20:157-160 [FREE Full text] [doi: 10.7326/M20-0912] [Medline: 32196071]
57. Henrich J, Heine SJ, Norenzayan A. Most people are not WEIRD. Nature 2010 Jul 01;466(7302):29. [doi: 10.1038/466029a] [Medline: 20595995]
58. Grady C. Payment of clinical research subjects. J Clin Invest 2005 Jul 01;115(7):1681-1687. [doi: 10.1172/jci25694]
59. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]

Abbreviations

RCT: randomized controlled trial
SAS: short and animated story-based

Edited by R Kukafka; submitted 15.04.21; peer-reviewed by B Nievas Soriano, S Badawy; comments to author 14.08.21; revised version received 08.09.21; accepted 12.09.21; published 22.11.21.

Please cite as:
Hachaturyan V, Adam M, Favaretti C, Greuel M, Gates J, Bärnighausen T, Vandormael A
Reactance to Social Authority in a Sugar Reduction Informational Video: Web-Based Randomized Controlled Trial of 4013 Participants
J Med Internet Res 2021;23(11):e29664
URL: https://www.jmir.org/2021/11/e29664
doi:10.2196/29664
PMID:34813490

Original Paper

Predictive Modeling of Vaccination Uptake in US Counties: A Machine Learning–Based Approach

Queena Cheong[1*]; Martin Au-yeung[2*]; Stephanie Quon[3]; Katsy Concepcion[2]; Jude Dzevela Kong[4], PhD

[1]School of Kinesiology, University of British Columbia, Vancouver, BC, Canada

[2]Faculty of Science, University of British Columbia, Vancouver, BC, Canada

[3]Faculty of Applied Science, University of British Columbia, Vancouver, BC, Canada

[4]Africa-Canada Artificial Intelligence and Data Innovation Consortium, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Jude Dzevela Kong, PhD
Africa-Canada Artificial Intelligence and Data Innovation Consortium
Department of Mathematics and Statistics
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100 ext 66093
Email: jdkong@yorku.ca

Abstract

Background: Although the COVID-19 pandemic has left an unprecedented impact worldwide, countries such as the United States have reported the most substantial incidence of COVID-19 cases worldwide. Within the United States, various sociodemographic factors have played a role in the creation of regional disparities. Regional disparities have resulted in the unequal spread of disease between US counties, underscoring the need for efficient and accurate predictive modeling strategies to inform public health officials and reduce the burden on health care systems. Furthermore, despite the widespread accessibility of COVID-19 vaccines across the United States, vaccination rates have become stagnant, necessitating predictive modeling to identify important factors impacting vaccination uptake.

Objective: This study aims to determine the association between sociodemographic factors and vaccine uptake across counties in the United States.

Methods: Sociodemographic data on fully vaccinated and unvaccinated individuals were sourced from several online databases such as the US Centers for Disease Control and Prevention and the US Census Bureau COVID-19 Site. Machine learning analysis was performed using XGBoost and sociodemographic data.

Results: Our model predicted COVID-19 vaccination uptake across US counties with 62% accuracy. In addition, it identified location, education, ethnicity, income, and household access to the internet as the most critical sociodemographic features in predicting vaccination uptake in US counties. Lastly, the model produced a choropleth demonstrating areas of low and high vaccination rates, which can be used by health care authorities in future pandemics to visualize and prioritize areas of low vaccination and design targeted vaccination campaigns.

Conclusions: Our study reveals that sociodemographic characteristics are predictors of vaccine uptake rates across counties in the United States and, if leveraged appropriately, can assist policy makers and public health officials to understand vaccine uptake rates and craft policies to improve them.

(J Med Internet Res 2021;23(11):e33231) doi:10.2196/33231

KEYWORDS

COVID-19; vaccine; public health; machine learning; XGBoost; SARS-CoV-2; sociodemographic factors; United States; sociodemographic; prediction; model; uptake

Introduction

The COVID-19 pandemic has affected millions worldwide. The widespread impact of the disease has forced populations into lockdown and self-isolation, and to social distance from each other to mitigate the disease spread. As a result, many await the successful development of a COVID-19 vaccine to return to normality. However, even if one becomes readily available, enough people need to have access or be willing to receive the vaccine to achieve herd immunity [1]. Previous literature has indicated disparities in vaccination rates between sociodemographic groups, and such factors play a substantial role in the likelihood of seeking vaccination. For example, those with lower education and income level [2,3], and Black individuals [4] are less likely to get vaccinated. Thus, the purpose of the study aims to use machine learning classification algorithms to construct a model that can predict vaccine uptake for US counties using publicly available sociodemographic data. Using this, public health officials can develop targeted interventions for specific populations to promote vaccine uptake by forecasting future vaccine behaviors. With the recent development in technological methods, researchers' use of machine learning methods to predict the likelihood of health behaviors has been on the rise. Previous studies have used XGBoost (extreme gradient boosting), a decision tree–based machine learning algorithm that uses a gradient-boosting framework, to build predictive models for vaccination uptake levels for influenza [5] and childhood immunizations [6]. Given the urgency of public health officials to encourage COVID-19 vaccination worldwide, such methods have substantial applicability in the current epidemiological context. Although many studies have investigated the impact of sociodemographic factors on vaccine uptake on a national level, research on a county level is scarce. Additionally, the use of data on smaller regions allows for a better understanding of local vaccine behaviors. This study seeks to fill this knowledge gap by incorporating a broad range of sociodemographic characteristics between US counties to predict COVID-19 vaccination uptake.

Methods

Sourcing the Data

Sociodemographic and socioeconomic data was collected from the US Department of Agriculture [7-9], the US Centers for Disease Control and Prevention [10], US Bureau for Labor Statistics [11], US Census Bureau COVID-19 Site [12-24], and Simple Maps US Counties Database [25]. In each database, data was collected for each US county and identified by its Federal Information Processing System (FIPS) code. This study included 2862 US counties out of 3007, and the counties were used based on the overlapping FIPS codes between data sets. The state of Texas was excluded as they did not release their COVID-19 vaccination data. The data set includes data collected between 2015 and 2021.

From these databases, 83 sociodemographic factors were collected and organized into 20 categories: education, ethnicity, income, employment, poverty, household size, population density, age, sex, disability status, access to technology, language spoken, health insurance, occupation, location, housing tenure, educational enrollment, grandparents taking care of grandchildren, access to income benefits, and working at home. Each category has between 1 and 14 subfactors. For example, the education category's factors are the percentage of adults with less than a high school diploma, percent of adults with a high school diploma only, percent of adults completing some college or associate degree, and percent of adults with a bachelor's degree or higher. A complete list of factors and their associated categories can be found in Table S1 in Multimedia Appendix 1. In addition, the percent of adults fully vaccinated against COVID-19 was found from the US Centers for Disease Control and Prevention [10]. The percentages are representative of vaccination data from May 21, 2021.

Creating a Universal Model

XGBoost Regressor was used as the predictive modeling algorithm to create a supervised regression model. XGBoost was chosen over other traditional machine learning methods because it is a decision tree–based model. This particular method can closely mimic the adaptive and consequential nature of the human decision-making process. In other words, a decision tree–based model can mimic how humans consider the potential outcomes of their actions before making a decision. Thus, our model provides a more accurate real-life depiction of how certain sociodemographic factors lead to decisions to take the vaccine. Furthermore, from an analytical standpoint, XGboost prevents overfitting and brings performance improvements compared to other traditional machine learning methods (eg, linear regression, elastic net, and random forest) since it uses a more regularized model formalization. In short, regularization is the process of adding information to solve a problem without overfitting, a process where a model fits too closely to its training data [26].

Other than performance alone, XGBoost has demonstrated great accuracy over other methods. For example, a previous study comparing the accuracy of different predictive modeling algorithms shows that XGBoost shows the highest accuracy score compared to other methods such as logistic regression, naive Bayes classifier, decision trees, and random forest [27]. Furthermore, XGBoost has demonstrated to learn better tree structures over decision tree models that use gradient boosting since XGBoost uses Newton boosting instead [28].

Lastly, we chose XGBoost because of its previous track record in the competitive machine learning scene. For example, in 2015, when Kaggle published the 29 winning solutions on their blog, it was found that 17 solutions used XGBoost [26]. The data science platform has also interviewed many of their top-ranking competitors on several occasions, and when asked what their favorite machine learning algorithms were, four members who have ranked as number one responded with XGBoost [28]. The annual data mining and knowledge discovery competition KDDCup 2015 further elucidates the system's prevalence, where the top 10 winning teams used XGBoost [26]. Examples of problems in these winning solutions include store sales prediction, ad click-through rate prediction, and hazard risk prediction [26]. The evident success of this method in solving myriads of real-life scenarios and problems

demonstrates its effectiveness and versatility in predictive health modeling.

In conclusion, with the aforementioned factors, XGBoost is a highly effective, efficient, and robust machine learning method with many benefits toward the needs of our paper.

Hypertuning Parameters for XGBoost

We used ExhaustiveGrid Search Cross-Validation (GridSearchCV) to perform five folds for each of the 384 permutations (totaling 1920 fits) to search for the optimal parameters to use in our XGBoost model to provide the highest accuracy in predicting vaccination uptake, specifically for our particular data sets. The learning rate represents how quickly an error is corrected from each tree. The max_depth determines the maximum depth a tree is allowed to grow during each boosting round. The min_child_weight is the minimum sum of instance weight needed in a child.

The subsample parameter randomly sets how much some of the training data is sampled prior to growing trees to prevent overfitting. The colsample_bytree parameter is the subsample ratio of columns when constructing each tree, again to prevent overfitting. N_estimators represent the number of trees to grow for the model. The range of parameters that we searched for the best fit is in Table S2 in Multimedia Appendix 1. The parameters we used in our model after computational fit (grid search best score of 0.5523) are illustrated in Table 1.

Table 1. Selected tuning parameters that were chosen after computational fit (grid search best score of 0.5523).

Parameters	Ranges
Learning rate	0.01
max_depth	9
min_child_weight	3
Subsample	0.7
colsample_bytree	0.7
n_estimators	1000

Evaluating the Model's Accuracy and Error

We adopted the use of k-fold cross-validation using the Scikit-learn package in Python (Python Software Foundation) to determine our accuracy score. A cross-validation method was chosen due to its ability to estimate the skill of a machine learning model based on unseen data. This can provide an estimate on how the model performs when used to make predictions on data not used during the training of our model—our accuracy percentage. The k-fold cross-validation uses a method where the cross-validation method is split into several groups that a given data sample is to be split into, defined by k. We chose a k value of 10, as this value is shown to have test error rate estimates that do not have high bias or high variance [29]. The final percentage accuracy representation produced by our k-fold cross-validation analysis is the percentage alignment with vaccination rates (percentage of population) in the test set with the predicted values.

We leveraged the root mean squared error (RMSE) to calculate our model's error, as RMSE is measured in the same units as the target variable (vaccination uptake percentage), providing us with a better interpretation and understanding of the errors of our models. RMSE measures the square root of the sum, for the vaccination uptake, of the square of the difference between the predicted (y_j) and actual vaccination uptake ($\hat{y}_j$), divided by the number of counties.

Evaluating a Feature's Importance

We used XGBoost's built-in feature importance, permutation analysis, and SHAP to understand how each feature drives our model's prediction score. Comparing these three aggregated methods allows us to understand how covariates contribute to the model fitting and the importance of the features we used for our vaccination model, rather than simply using one method. To avoid bias due to specific feature categories containing more factors, we used Breiman's permutation-based measures to assess the importance of a feature by calculating the degree of increase in the model's prediction error after their values are permuted, otherwise known as randomized ablation [30]. This approach provided us with a way to investigate the independent predictive power of each category without building separate machine learning models for each feature category to do so. We performed permutation analysis using the Python Scikit-learn library. The SHAP library provided a game-theoretic approach to determine an overview of which features are most important for a model by plotting the absolute mean SHAP value in a bar graph [31]. Lastly, XGBoost's built-in feature importance calculated how important a feature is from its corresponding score. We then aggregated the top five features from these three methods; then, we selected all features demonstrated in the three feature importance analysis methods as our final list of selected important features driving vaccination uptake.

Results

Using XGBoost's regressor function hypertuned with custom parameters, our model predicted the percentage of COVID-19 vaccination uptake per county with an accuracy of 62%. This accuracy score was calculated based on the k-fold cross-validation average score.

Our model demonstrated relatively low errors using RMSE analysis, where it demonstrated a 0.08% vaccination uptake percentage error compared to the test model and 0.05%

vaccination uptake percentage error compared to the actual model.

The choropleth map of US counties shown in Figure 1 demonstrates our machine learning model's predicted percentage of vaccine uptake. White areas on the choropleth maps represent areas with no vaccination uptake data or other missing sociodemographic data.

In Figure 2, the choropleth map of US counties shows the difference between actual uptake and predicted uptake, and highlights counties where our model was less accurate in predicting vaccination uptake.

Figure 1. Predicted vaccination uptake percentage by US counties. White areas represent areas with no vaccination uptake data.

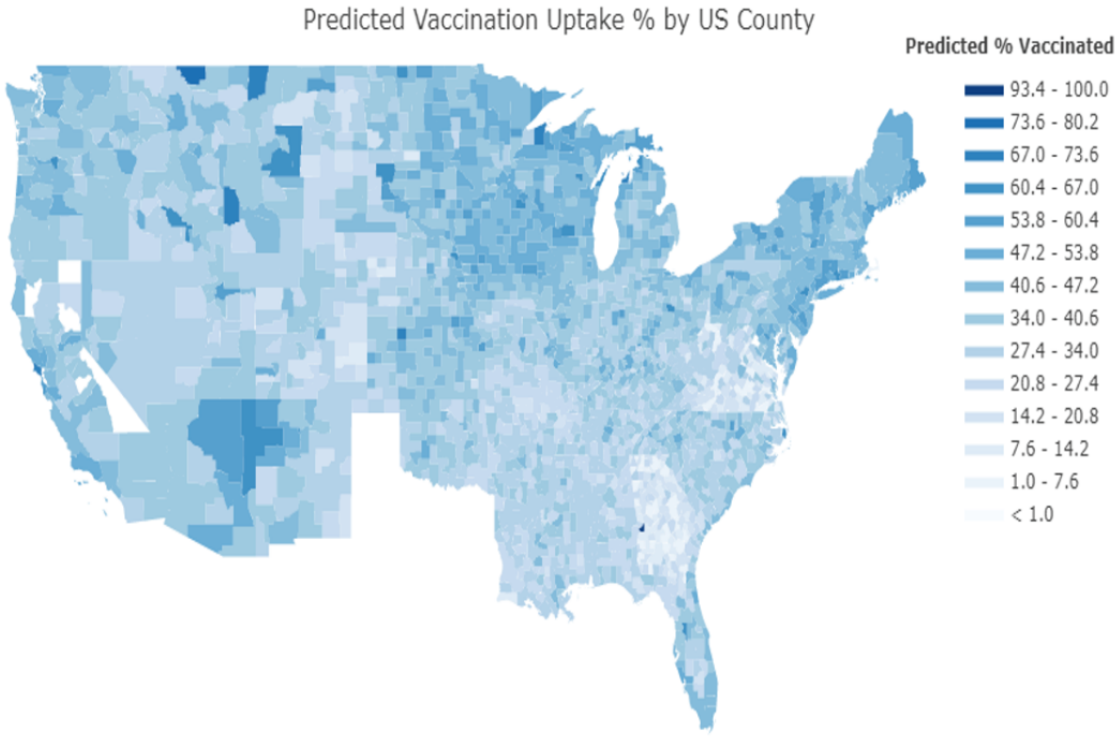


Figure 2. Accuracy of the model in predicting actual vaccination uptake by US county. The darkest shades of red represent lower prediction accuracy, with white representing the highest prediction accuracy.

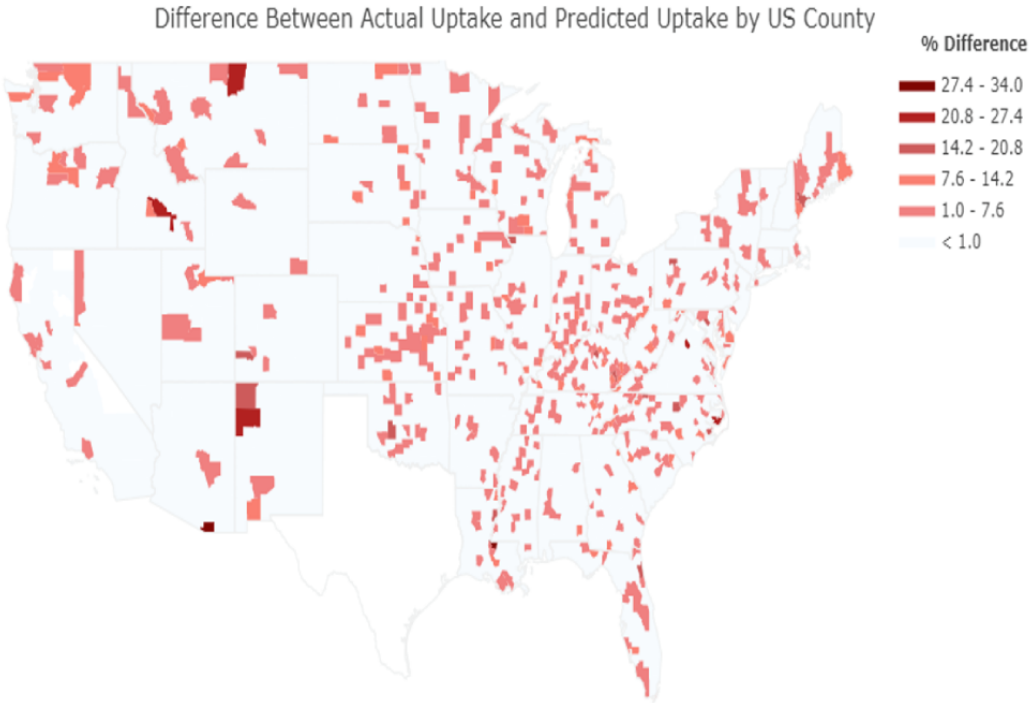


Figure 3 highlights the top five features generated from XGBoost's built-in plot importance function in ranking order of importance. The top five features that were identified to drive XGboost's predictive model were geographic location (longitude, latitude), adults with less education (percent of adults with a high school diploma only), indigenous population (percent non-Hispanic American Indian/Alaska Native), and median household income (median household income percent of the state). The top 25 features generated from XGBoost's built-in plot importance can be found in Table S3 in Multimedia Appendix 1.

Figure 4 illustrates the top five features generated by the Python Scikit-learn package permutation importance function in ranking order of importance. The most notable features found are geographic location (longitude, latitude), education (percent of adults with less than a high school diploma), online access (households with broadband internet), and income (median household income percent of the state). The top 25 features generated by this approach can be found in Table S4 in Multimedia Appendix 1.

Figure 5 demonstrates the top five features generated by SHAP in ranking order of importance. The top significant features that SHAP found to drive our predictive model based on SHAP were geographic location (longitude, latitude), education level (percent of adults with less than a high school diploma, percent of adults with a bachelor's or higher), and online access (households with broadband internet). The ranking influence of the remaining features generated by this approach can be found in Figure S2 in Multimedia Appendix 1.

Figure 3. The top five identified sociodemographic factors to predict vaccination uptake by XGBoost's built-in feature importance analysis function.

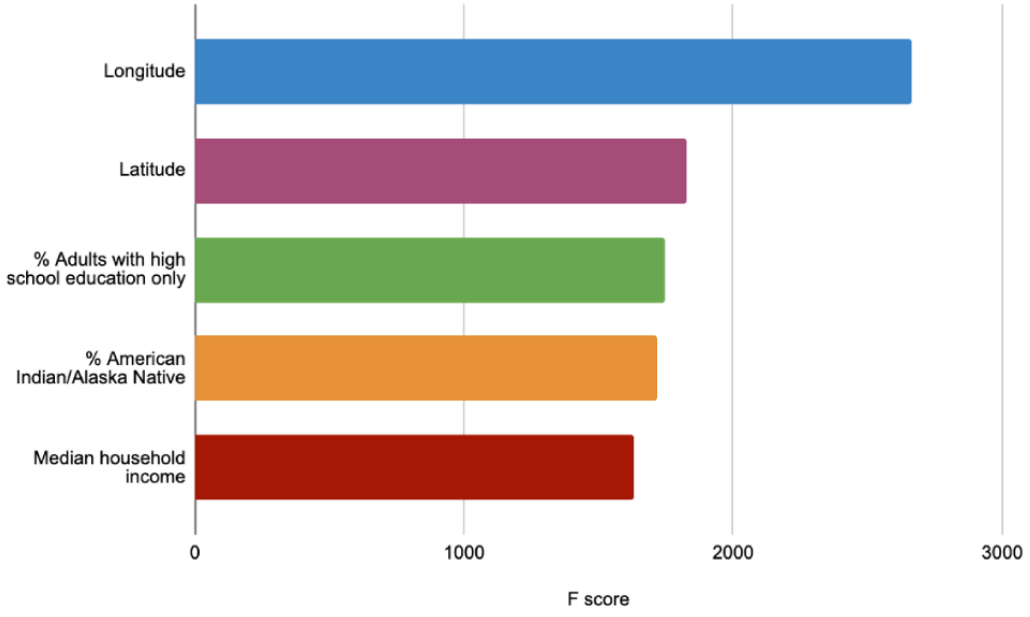


Figure 4. The top five identified sociodemographic factors to predict vaccination uptake found by permutation analysis.

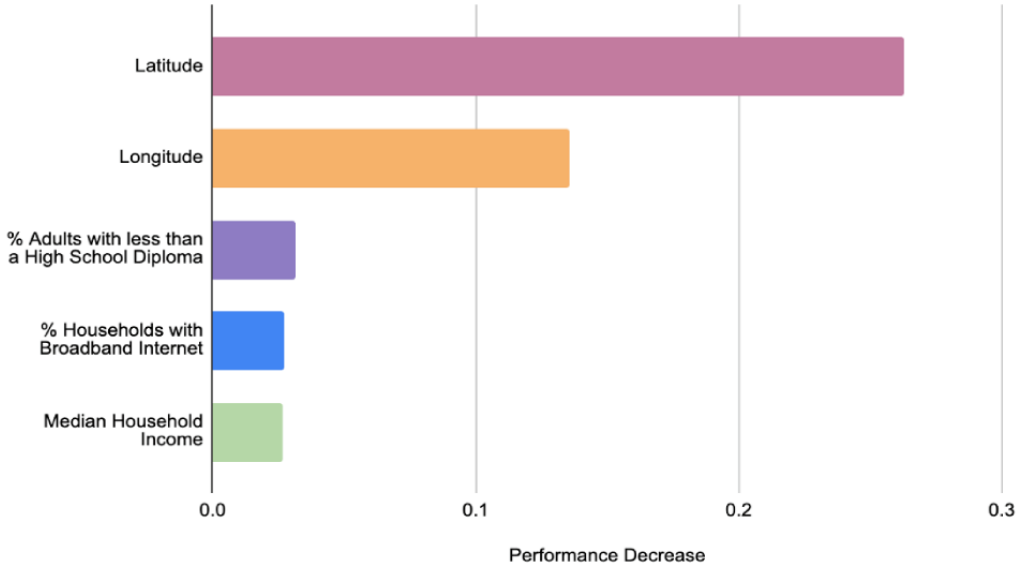
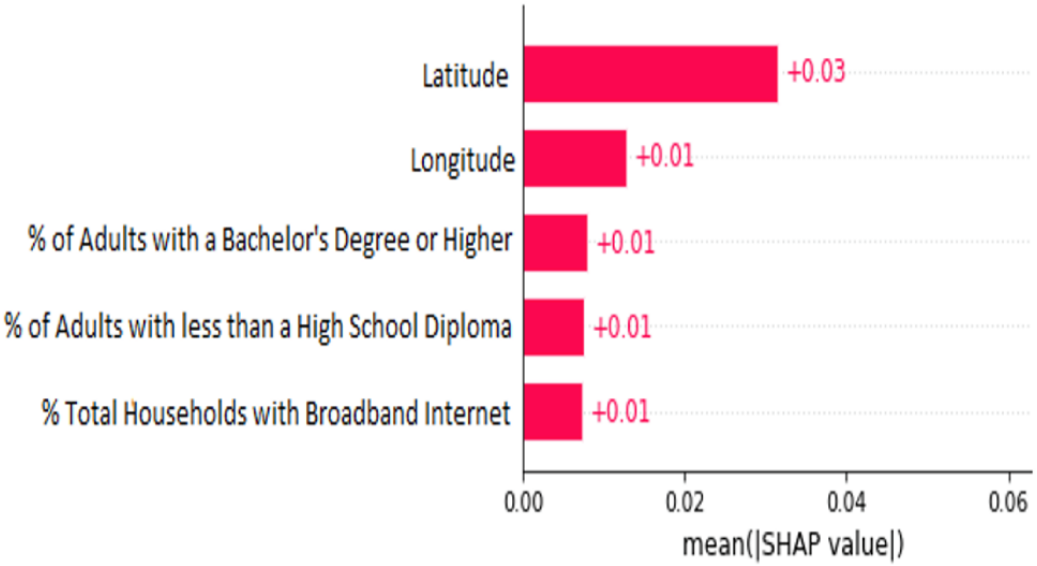


Figure 5. The top five identified sociodemographic factors that drive the model's prediction for vaccination uptake found by SHAP.



Discussion

Our XGBoost model scored a 62% accuracy score in predicting vaccination uptake based on 83 sociodemographic factors in this study. We also determined that geographic location, education, household accessibility to broadband internet, median household income, and ethnicity were the six main factors driving our model's prediction across analyses done by XGBoost, permutation analysis, and SHAP.

Accuracy of Predictive Modeling

Our machine learning model scored a k-fold cross-validation accuracy score of 62%, representing the model's ability to accurately predict vaccination uptake based on sociodemographic factors. The accuracy score alone does not provide sufficient information regarding the mechanism by which a sociodemographic factor impacts a population. However, the machine learning model can produce a choropleth with individualized percentage accuracy scores per county, where health authorities and governing bodies can use these percentage scores to visualize areas with lower than average vaccination uptake. Therefore, the results of this machine learning analysis can advise those in charge of mitigating the pandemic in identifying at-risk areas in need of targeted vaccination campaigns.

In Figure S1 in Multimedia Appendix 1, we plotted the actual vaccine uptake per county on a choropleth of the United States. Darker regions represent areas with higher vaccination rates, while lighter regions represent areas with lower vaccination rates. Figure 1 shows our machine learning model's predicted vaccination rate per county. When comparing the two plotted maps, the observed differences between the actual and predicted models were less distinct, denoting evidence of our model's ability to visualize vaccination uptake accurately. The state of Texas was excluded from our map, as they did not release their COVID-19 vaccination uptake data.

However, our current model does not consider additional nonsociodemographic factors such as policies enacted by local governments, political views in local areas, and citizens' general behavior, limiting our model's accuracy score. Therefore, to encompass vaccination rate and increase the accuracy score of our model, we must also view nonsociodemographic factors, as they can also substantially drive an individual's decision to take the vaccine.

To further evaluate the accuracy of our regression model, we generated a choropleth identifying the counties where our model had difficulties or ease in predicting vaccination (Figure 2). As evident, there are few counties where the model had low prediction accuracy—the worst being a 27.4% to 34.0% difference between the predicted and actual vaccination uptake. The two counties where the regression model struggled with predicting the most are Santa Cruz, Arizona (AZ) and West Feliciana, Los Angeles (LA). As well, other counties that the model yielded a 20.8% to 27.4% difference between predicted and actual vaccination uptake are McKinley, New Mexico (NM); Blaine County, Montana (MT); and Blaine County, Idaho (ID).

The generated prediction accuracy choropleth in Figure 2 can be helpful for public health authorities since it identifies which counties the model was accurate or inaccurate in predicting actual vaccination uptake. With this knowledge, officials and health governing bodies can better understand and decide which areas they need to target their efforts toward. In addition, the counties that our analysis has identified to have low prediction accuracies, such as Santa Cruz, AZ, can be an avenue for future research to investigate why there was a high prediction error in those particular counties.

As well, the predictability accuracy can allow public health authorities in future pandemics to predict more precisely an estimated vaccination uptake based solely on sociodemographic factors. Thus, targeting ahead of time areas that should prioritize education on vaccination safety and why it is essential to receive the vaccine.

Feature Importance

The significance of the top identified sociodemographic features by our model—location, education, ethnicity, income, and access

to the internet—provide a vivid portrayal of the current social climate in the United States and is calculated by using XGBoost's built-in feature importance (Figure 3), permutation analysis (Figure 4), and SHAP (Figure 5).

Based on the following three methods in determining feature importance, we determined that geographic location has the most influence on our prediction model, with both latitude and longitude ranking first or second, respectively. As well, the third feature is primarily dominated by an educational-based factor (school level).

Our feature importance analyses ranking location and education so highly also provides a more comprehensive look at other aspects of American features that are prominent presently, such as lack of education and political divide in rural areas compared to more populated areas. Our evidence supports that these sociodemographic factors significantly influence disparities in access to health resources and must continue to be the focus of public and government efforts to decrease the gap.

Longitude and Latitude

Based on our feature importance analysis methods, our machine learning model determined that longitude and latitude were the top two most crucial factors across all three methods, suggesting that geographic location plays the most prominent role in vaccine uptake. There may be an interaction term between longitude and latitude, given there may be a three-way relationship between longitude, latitude, and vaccine uptake. In past research studies, individuals residing in the western United States were more likely to refuse their children's vaccinations [32], and a recent study found that COVID-19 vaccination coverage is lower in rural counties [33]. In addition, population density, which ties in closely with the significance of a geographic region, ranked ninth in importance on our factors list. With this in mind, our results demonstrate that geographic location has a clear role in driving vaccination uptake. However, our results do not precisely determine the direct relationship between how geographic location influences vaccine uptake. Therefore, future research could explore the specific implications of location and living in rural areas to determine where additional COVID-19 vaccination centers can be opened.

Education

Although this study explored several educational factors related to level of education and educational enrollment, the factor with the highest rank of importance was the percent of adults with a high school diploma only. This factor ranked third in importance, and two other educational factors ranked in the top 8 factors.

Previous research has indicated a relationship between vaccine uptake and education. In a study, individuals who had attained higher levels of education were more likely to accept vaccines as safe [34] and vaccinate individuals in their care, such as their children [35]. As our results do not determine the direct relationship between education and vaccine uptake, it cannot be concluded that level of education increases or decreases vaccine uptake. However, our results demonstrate that educational groups separated by secondary school attainment could be an important factor in determining whether an individual will receive the COVID-19 vaccine. Therefore, future research could explore this specific factor and how it impacts vaccine uptake.

Ethnicity

Despite being twice as likely to die from COVID-19 [36], there is consistent evidence that ethnic minority groups are less likely to be vaccinated for the virus (eg, [37]) compared to their White counterparts. Factors that have been explained to drive this inequality include differences in vaccine hesitancy [38], attitudes toward vaccines [39], and trust in distributing pharmaceutical companies between ethnic groups [39]. In our findings, ethnicity is the fourth most important factor associated with vaccination uptake, providing evidence that an individual's cultural background is substantially associated with whether or not they receive vaccination. Thus, these findings provide support that there is a need for the development of special efforts to target historically marginalized populations in vaccination campaigns and increase vaccination rates among those groups with low uptake.

Median Household Income

In previous studies, individuals from lower-income households were less confident in the importance of vaccines [40] while also being more vulnerable to the impacts of COVID-19 [35]. Household income also relates closely to socioeconomic status, and individuals belonging to higher socioeconomic status groups are more likely to receive vaccines [41]. This combination of factors stemming from income and financial status may impact an individual's ability and willingness to receive the COVID-19 vaccine. Our finding that household income is an important factor in vaccine uptake is consistent with previous literature. However, our results do not determine if lower or higher income households are more likely to receive the COVID-19 vaccine. Therefore, more research is needed to explore the relationship between these two factors. However, given the previous evidence that individuals from lower-income groups are less likely to receive vaccinations, vaccine campaigns could further target their efforts toward lower-income groups to ensure that more individuals have access to vaccines and have confidence in them.

Internet Accessibility

Our feature importance analyses also revealed that household access to broadband internet was significant in predicting vaccination uptake. This is an important find, as digital technology played a substantial role during the COVID-19 pandemic in communicating health information from administrations to the public, aiding disease surveillance, and developing mobile health apps [42]. In addition, the convenience of social media enables many communities, particularly historically marginalized groups, to access critical COVID-19 data and information more readily and easily. Thus, many public health agencies sought online appointment platforms to assist with their vaccination booking processes during the pandemic. This means that having direct access to the internet can play a role in determining whether an individual can receive a vaccine. However, being knowledgeable and literate on how to use it may possibly be even more pivotal.

Published studies have reported racial and educational differences in digital literacy. For example, a US Department of Education report posits that Black people were twice as likely to be digitally illiterate than their White counterparts [43]. An individual's level of formal education also affects their knowledge of computer literacy [44]. Such factors may contribute heavily to the ability of individuals to identify misinformation on the internet and the desire to get a COVID-19 vaccination—otherwise known as the degree that they are vaccine hesitant. The World Health Organization has cited vaccine hesitancy as one of the top 10 threats to global health, as this delay in acceptance threatens the process of tackling widespread viruses and diseases [45].

With the study's finding that asserts that household access to the internet is primary in predicting one's vaccination uptake, we can bring forth awareness to public health officials about the importance of centering their efforts in providing greater accessibility to broadband internet in communities that may not have widespread internet use and teach them about digital literacy. This will enable those communities with the skills to critically interpret the vast proliferation of health information during this pandemic. Doing so could potentially help certain groups alleviate vaccine concerns, better understand the scientific rationale behind vaccines, and recognize misinformation when they encounter it.

Conclusions

Although in the United States, COVID-19 cases are moving toward a downward trajectory and counties are beginning to fully reopen, the study is important for future pandemics or even if new variants may require new vaccination development. Furthermore, by constructing a model that can predict future vaccine behaviors in US counties, we can better advise public health authorities in advance, allowing them to prepare areas of vaccination campaign focus more efficiently and effectively.

Acknowledgments

We would like to thank the STEM Fellowship for organizing a data science challenge.

QC would like to dedicate this paper to her doctor grandfather, who studied at Beijing's "third top medical school." He absolutely loved spending his time annotating every word and sentence in English magazines to learn the language, spending all day at the library reading, and improving his medical knowledge by reading English research papers. He would have had a field day at the library deciphering this one.

MA would like to dedicate this paper to his parents, Eliza Wong and Alan Au-yeung. This paper would not be possible if they did not give MA permission to purchase a new computer that could run the machine learning models (RTX 3070).

JDK acknowledges support from Canada's International Development Research Centre and Swedish International Development Cooperation Agency (grant 109559-001).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of their institutions.

All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[DOCX File , 175 KB - jmir_v23i11e33231_app1.docx]

References

1. Griffith J, Marani H, Monkman H. COVID-19 vaccine hesitancy in Canada: content analysis of tweets using the theoretical domains framework. J Med Internet Res 2021 Apr 13;23(4):e26874 [FREE Full text] [doi: 10.2196/26874] [Medline: 33769946]
2. Callaghan T, Moghtaderi A, Lueck JA, Hotez P, Strych U, Dor A, et al. Correlates and disparities of intention to vaccinate against COVID-19. Soc Sci Med 2021 Mar;272:113638 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113638] [Medline: 33414032]
3. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes toward a potential SARS-CoV-2 vaccine: a survey of U.S. adults. Ann Intern Med 2020 Dec 15;173(12):964-973 [FREE Full text] [doi: 10.7326/M20-3569] [Medline: 32886525]
4. Uscher-Pines L, Maurer J, Harris KM. Racial and ethnic disparities in uptake and location of vaccination for 2009-H1N1 and seasonal influenza. Am J Public Health 2011 Jul;101(7):1252-1255 [FREE Full text] [doi: 10.2105/AJPH.2011.300133] [Medline: 21566026]

5. Shaham A, Chodick G, Shalev V, Yamin D. Personal and social patterns predict influenza vaccination decision. BMC Public Health 2020 Feb 12;20(1):222 [FREE Full text] [doi: 10.1186/s12889-020-8327-3] [Medline: 32050948]
6. Mannion N. Predictions of changes in child immunization rates using an automated approach: USA. NORMA@NCl Library. 2020 Sep 28. URL: http://norma.ncirl.ie/4368/1/niallmannion.pdf [accessed 2021-11-21]
7. Educational attainment for the U.S., states, and counties, 1970-2019. US Department of Agriculture Economic Research Service. URL: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data [accessed 2021-11-19]
8. Unemployment and median household income for the U.S., States, and counties, 2000-20. US Department of Agriculture Economic Research Service. 2021 Jun 02. URL: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/ [accessed 2021-11-19]
9. Poverty estimates for the U.S., States, and counties, 2019. US Department of Agriculture Economic Research Service. URL: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/ [accessed 2021-11-19]
10. Centers for Disease Control and Preventio. Fully vaccinated adults. Centers for Disease Control and Prevention. URL: https://data.cdc.gov/Vaccinations/Fully-Vaccinated-Adults/jm79-dz78 [accessed 2021-11-19]
11. Labor force data by county, 2019 annual averages. US Bureau of Labor Statistics. 2019. URL: https://www.bls.gov/LAU/ [accessed 2021-11-19]
12. Population and poverty status - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/d8f418b6357841d6af70d511372d8aac_2/explore?location=36.005565%2C-96.362692%2C3.80 [accessed 2021-11-19]
13. Average household size and population density - counties. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/21843f238cbb46b08615fc53e19e0daf/explore?location=11.971969%2C0.315550%2C1.74 [accessed 2021-11-19]
14. Population by age and sex - counties. US Census Bureau COVID-19 Site. 2021 Nov 08. URL: https://covid19.census.gov/datasets/407c6b6b567947928b6d6cb6e76b85a9_2/explore?location=14.610835 [accessed 2021-11-19]
15. Disability status of the civilian noninstitutionalized population - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/c76d5b9a55ac4c659b74670b11d7a3c6_2/explore?location=18.834312 [accessed 2021-11-19]
16. Computers and internet use - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/190e48c808e0457e90a5b92086a41312_2/explore?location=14.610835 [accessed 2021-11-19]
17. Language is spoken at home - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/2cd2ac0e318e4e678c33ebf0c7ad37ec_2/explore?location=14.610835 [accessed 2021-11-19]
18. Health insurance coverage - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/1420ea3c03cf44f7a4872c4ff8876b9a_2/explore [accessed 2021-11-19]
19. Occupation by sex - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/6e21c8eca95d4b39a5d55e2f7ad5ed92_2/explore?location=19.320508 [accessed 2021-11-19]
20. Housing tenure - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/2817f261a24b4d63b0299e7d7faf7997_2/explore?location=5.519913 [accessed 2021-11-19]
21. School enrollment - counties, 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/4011262b4210470eb2ee5d7f4dd6f335_2/explore?showTable=true [accessed 2021-11-19]
22. Grandparents - counties, 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/1364684f2f88400c8d96cb6c7cb913c4_2/explore [accessed 2021-11-19]
23. Income and benefits - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/fa2c4f34c4b24d6f87bc2a2fbdd1220f_2/explore?location=17.051452 [accessed 2021-11-19]
24. Worked at home - counties 2015-2019. US Census Bureau COVID-19 Site. URL: https://covid19.census.gov/datasets/a50a9052f46a4f5d9775336f76405ba2_2/explore?location=16.891069 [accessed 2021-11-19]
25. United States counties database. Simple Maps. URL: https://simplemaps.com/data/us-counties [accessed 2021-11-19]
26. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Aug 13 Presented at: KDD '16; August 13-17, 2016; San Francisco, CA p. 785-794.
27. Abdurrahim Y, Ali Anil D, Sena K, Huseyin U. Comparison of deep learning and traditional machine learning techniques for classification of pap smear images. arXiv. Preprint posted online on September 11, 2020 [FREE Full text]
28. Nielsen D. Tree boosting with XGBoost: why does XGBoost win "every" machine learning competition? NTNU Open. 2016. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2433761/16128_FULLTEXT.pdf [accessed 2021-11-19]
29. Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer; Sep 01, 2013.
30. Breiman L. Random forests. Machine Learning 2001 Oct 01;45:5-32.
31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Dec 04 Presented at: 31st Conference on Neural Information Processing Systems; 2017; Long Beach, CA URL: https://arxiv.org/abs/1705.07874/
32. Gust DA, Darling N, Kennedy A, Schwartz B. Parents with doubts about vaccines: which vaccines and reasons why. Pediatrics 2008 Oct;122(4):718-725. [doi: 10.1542/peds.2007-0538] [Medline: 18829793]

33. Murthy BP, Sterrett N, Weller D, Zell E, Reynolds L, Toblin RL, et al. Disparities in COVID-19 vaccination coverage between urban and rural counties - United States, December 14, 2020-April 10, 2021. MMWR Morb Mortal Wkly Rep 2021 May 21;70(20):759-764. [doi: 10.15585/mmwr.mm7020e3] [Medline: 34014911]
34. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
35. Viswanath K, Bekalu M, Dhawan D, Pinnamaneni R, Lang J, McLoud R. Individual and social determinants of COVID-19 vaccine uptake. BMC Public Health 2021 Apr 28;21(1):818 [FREE Full text] [doi: 10.1186/s12889-021-10862-1] [Medline: 33910558]
36. Risk for COVID-19 infection, hospitalization, and death by race/ethnicity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html [accessed 2021-11-19]
37. Iacobucci G. Covid:19: ethnic minority health staff are less likely to take up vaccine, early data show. BMJ 2021 Feb 16;372:n460. [doi: 10.1136/bmj.n460] [Medline: 33593814]
38. Khan MS, Ali SAM, Adelaine A, Karan A. Rethinking vaccine hesitancy among minority groups. Lancet 2021 May 22;397(10288):1863-1865 [FREE Full text] [doi: 10.1016/S0140-6736(21)00938-7] [Medline: 33894143]
39. Grumbach K, Judson T, Desai M, Jain V, Lindan C, Doernberg SB, et al. Association of race/ethnicity with likeliness of COVID-19 vaccine uptake among health workers and the general population in the San Francisco Bay Area. JAMA Intern Med 2021 Jul 01;181(7):1008-1011 [FREE Full text] [doi: 10.1001/jamainternmed.2021.1445] [Medline: 33783471]
40. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet 2020 Sep 26;396(10255):898-908 [FREE Full text] [doi: 10.1016/S0140-6736(20)31558-0] [Medline: 32919524]
41. Lackner CL, Wang CH. Demographic, psychological, and experiential correlates of SARS-CoV-2 vaccination intentions in a sample of Canadian families. Vaccine X 2021 Aug;8:100091 [FREE Full text] [doi: 10.1016/j.jvacx.2021.100091] [Medline: 33778480]
42. The Lancet Digital Health. Can technology increase COVID-19 vaccination rates? Lancet Digit Health 2021 May;3(5):e274 [FREE Full text] [doi: 10.1016/S2589-7500(21)00061-3] [Medline: 33814326]
43. Mamedova S, Pawlowski E. A description of U.S. adults who are not digitally literate. National Center for Education Statistics. URL: https://nces.ed.gov/pubs2018/2018161.pdf [accessed 2021-11-19]
44. Milić M, Škorić I. The impact of formal education on computer literacy. Int J Emerging Technologies Learning 2010 Mar 18;5(SI2):60. [doi: 10.3991/ijet.v5s2.1253]
45. Ten threats to global health in 2019. World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-11-19]

Abbreviations

FIPS: Federal Information Processing System
RMSE: root mean squared error

Edited by C Basch; submitted 28.08.21; peer-reviewed by S Ghaemi, M Che; comments to author 17.09.21; revised version received 17.10.21; accepted 01.11.21; published 25.11.21.

Please cite as:
Cheong Q, Au-yeung M, Quon S, Concepcion K, Kong JD
Predictive Modeling of Vaccination Uptake in US Counties: A Machine Learning–Based Approach
J Med Internet Res 2021;23(11):e33231
URL: https://www.jmir.org/2021/11/e33231
doi:10.2196/33231
PMID:34751650

Original Paper

Using Ecological Momentary Assessment to Study the Development of COVID-19 Worries in Sweden: Longitudinal Study

Peter Johannes Schulz[1], PhD; Elin M Andersson[2], MSc; Nicole Bizzotto[1], MA; Margareta Norberg[3], MD, PhD

[1]Institute of Communication and Health, Università della Svizzera italiana, Lugano, Switzerland

[2]Department of Psychology, Umeå University, Umeå, Sweden

[3]Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden

Corresponding Author:
Peter Johannes Schulz, PhD
Institute of Communication and Health
Università della Svizzera italiana
Via Giuseppe Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58666 ext 4724
Email: schulzp@usi.ch

Abstract

Background: The foray of COVID-19 around the globe has certainly instigated worries in many people, and lockdown measures may well have triggered more specific worries. Sweden, more than other countries, relied on voluntary measures to fight the pandemic. This provides a particularly interesting context to assess people's reactions to the threat of the pandemic.

Objective: The general aim of this study was to better understand the worried reactions to the virus and the associated lockdown measures. As there have been very few longitudinal studies in this area published to date, development of feelings of worry over time was analyzed over a longer range than in previous research. Affective variables, worry in particular, were included because most of the research in this field has focused on cognitive variables. To employ new methodology, ecological momentary assessment was used for data collection and a multilevel modeling approach was adopted for data analysis.

Methods: Results were based on an unbalanced panel sample of 260 Swedish participants filling in 3226 interview questionnaires by smartphone over a 7-week period in 2020 during the rapid rise of cases in the early phase of the pandemic. Causal factors considered in this study included the perceived severity of an infection, susceptibility of a person to the threat posed by the virus, perceived efficacy of safeguarding measures, and assessment of government action against the spread of COVID-19. The effect of these factors on worries was traced in two analytical steps: the effects at the beginning of the study and the effect on the trend during the study.

Results: The level of general worry related to COVID-19 was modest (mean 6.67, SD 2.54 on an 11-point Likert scale); the increase during the study period was small, but the interindividual variation of both the worry level and its increase over time was large. Findings confirmed that the hypothesized causal factors (severity of infection, susceptibility to the threat of the virus, efficacy of safeguarding, and assessment of government preventive action) did indeed affect the level of worry.

Conclusions: The results confirmed earlier research in a very special case and demonstrated the usefulness of a different study design, which takes a longitudinal perspective, and a new type of data analysis borrowed from multilevel study design.

(J Med Internet Res 2021;23(11):e26743) doi:10.2196/26743

KEYWORDS

COVID-19; coronavirus; longitudinal studies; EMA; worry; fear; pandemics

Introduction

When a serious health threat approaches, people have to decide whether any protection measures are called for. When it comes to explaining or predicting self-protective behavior as a response to the threat of a communicable viral infection, theorizing and empirical results related to cognitive predictors are available in abundance. These are most prominently represented by the cognitive factors of threat severity [1] and perceived personal susceptibility [2] in the process of impression formation, the

efficacy of measures against the threat (response efficacy [3]), and one's ability to follow these measures (self-efficacy). Several conceptualizations and theories stand at hand to show a comprehensive picture of how all of these aspects fit together [4-10].

Besides these cognitive predictors, there are also elements of affective or emotional factors of self-protective behavior, which are treated together with cognitive factors in some cases and are treated separately in other cases. For instance, Harper and colleagues [11] summarize that research has paid substantial attention to public communication, but only little attention has been paid to character and emotion as factors of protection behavior. However, the literature on affective reactions in pandemics seems to be mainly concerned with the consequences rather than with the predictors of worry. In a study on the responses of a Chinese population to the pandemic, MacKay et al [12] reached a similar conclusion for the "routes to anxiety over disease contraction," which they refer to as "understudied." Assuming theoretical positions, Presti et al [13] argue that fear is an integral part of pandemics, which is mostly damaging but can also be contained (see also [14]). Early results from application of the Fear of COVID-19 Scale (FCV-192), a questionnaire designed for use worldwide, bear witness to this fact (eg, [15]).

Evidence of strong affective reactions to COVID-19 among the public is highlighted by the calls to action raised by Asmundson and Taylor [16,17], who refer to the first survey studies in connection with the global COVID-19 crisis. Several researchers have aimed to interpret the consequences of COVID-19 clinically and presented them as mental conditions [18,19], compared anxiety related to the present virus with the health anxiety trait measured several months earlier, and linked this anxiety with "cyberchondria," an exaggerated need to seeking health information [20]. The concept of a behavioral immune system [12] was tested and largely confirmed, supporting the assumption that individuals take actions that involuntarily protect them from infection at times of a pandemic [21].

If fear indeed takes this role, this could explain the acceptance of restrictions in the fight against the virus that was found in a study from Denmark [22]. In a British study, Harper et al [11] assessed the self-perceived risk of contracting COVID-19, fear of the virus, moral rules, and political ideology on behavior change in response to the pandemic. Fear of COVID-19 turned out to be the only predictor of adopting protective measures [11]. In a Turkish study, the very same variable, fear of COVID-19, was shown to affect several mental variables [23]. These findings provide sufficient reason to perform more research on the potential effects of trait and situational emotional states, and in particular of worries and fears of communicable viral infections, on behavioral intentions for protecting oneself.

Worries and fear are not only important as a predictors of behavior but it is also relevant to ask how emotions develop and to identify their predictors. A longitudinally designed Chinese study of emotional reactions to fear of communicable diseases, severe acute respiratory syndrome (SARS) in this case, found that older and middle-aged people experienced less anger and had less need for emotion-focused coping skills in comparison with younger adults. Over the complete study period, emotion-focused coping increased more among the older and middle-aged population than among the younger participants; however, this trend was reversed at the peak of the SARS outbreak. The overall age differences were then reversed by the end of the outbreak. Findings of this study suggest that older adults may be better at emotional regulation than their younger counterparts: they react to a crisis with less anger and are better able to adapt their coping strategies to the changing environment [24]. This volatility of coping abilities motivated us to focus on worry, as a disagreeable and often uncontrollable state, in our present analysis. A review of demographic and attitudinal determinants of protective behaviors during a pandemic showed that being a woman, having a higher educational level, and being older are associated with behavior modification. Additionally, individuals' perceived susceptibility to and severity of the disease, as well as stronger belief in the effectiveness of recommended safeguarding behaviors predict behavior change. Moreover, trust in authorities and a higher level of anxiety were also associated with compliance with protective behaviors [25].

Assuming that emotions affect behavior, identifying the causal influences on emotions is an eminent question. Focusing on three serious mental conditions as dependent variables (depression, anxiety disorders, and posttraumatic stress disorder), a study performed in Spain found that the elderly, the well-off, and people who felt adequately informed of COVID-19 developed these mental conditions less often, whereas women, people with a history of mental conditions, those with present COVID-19 symptomatology, and those with experience of others close to them having COVID-19 had increased mental health symptoms. Spiritual well-being was the best predictor for avoiding mental health symptoms and loneliness was the highest risk factor [18]. The measures taken against the spread of the pandemic also have to be considered as a separate source for adverse mental states [26].

There are few studies of the COVID-19 pandemic that assumed a longitudinal perspective. A German survey study showed that participants reported significantly increasing virus-related anxiety in the months before the survey. As these months were over when the survey was fielded, the information about the trend development had to be collected retrospectively. Cyberchondria in the pandemic was associated with current virus anxiety, and the association was moderated by the trait health anxiety. Subjectively adequate information on the virus lowered current virus anxiety [27].

A Chinese study of the mental health and affective consequences of confinement shortly after the virus appeared in Wuhan also adopted a longitudinal perspective by repeating measures from before confinement to 2 weeks into confinement. Such an approach may be adequate for studying the consequences of confinement on those who had to suffer from it, but does not suffice as a major step in assessing the time series on macrosocial responses to the pandemic [28]. As an aside, such research suggests, as do other experts in the field [13], that people should be put in quarantine only after benefits have been weighted against risks.

Based on these considerations, the aim of this study was to describe the development of Swedish people's reaction to the spread of the COVID-19 virus (SARS-CoV-2), measured predominately as general worries related to the pandemic. The observations from previous studies highlighted above led to some conclusions for our study. As cognitive predictors of self-protective behavior have so far been studied more broadly than affective predictors, we were more interested in the latter, and finally chose a design that contained predictors of both types. The dynamic nature of affective reactions and fears demands studies with a longitudinal design; yet, such designs have hardly ever been employed in studies on the subject. To fill this gap, we used a longitudinal design with data from a range suitable to describe change over time during the very early phase of the COVID-19 pandemic. We also employed newly developed methods for data collection and data analysis. In a longitudinal design with high-frequency measures of 1 day, we employed an ecological momentary assessment (EMA) design for data collection. EMA has been discussed and successfully used for a variety of subjects [29-32]. For data analysis, a multilevel modeling approach was employed.

During the early phase of the COVID-19 pandemic, government assessment, communication, and actions taken highly differed between countries. Although many governments decided to lock down large parts of society in an attempt to curb the spread of the virus, the Swedish society, by contrast, was not closed, but safeguarding measures were launched, and the population was urged to voluntarily follow recommendations similar to a lockdown. Relying on voluntary recommendations gives the population's perceptions and opinions of the threat created by the pandemic a special importance. At the same time, the expectations from the government and other authorities on individuals to show solidarity, take responsibility, and follow recommendations and regulations were strong in Sweden [33]. Some formal restrictions such as prohibition of visits to homes for the elderly and rules for distancing at restaurants were also implemented. During this period, the number of deaths increased dramatically and reached considerably higher levels compared to, for example, the surrounding Nordic countries. The Swedish government's policy was widely debated and strongly questioned. For example, it was suggested in national and international media that the population was exposed to an "experiment." However, by the end of April 2020, Sweden's policy was also supported by the World Health Organization as a "role model" [34].

The aim of this study was to assess worries of COVID-19, predictors of these worries, as well as individuals' evaluations of the lockdown measures in Sweden. Moreover, as there have so far been very few longitudinal studies in this research area, the aim was to study development over time for a longer range than adopted in previous research. Special attention was paid to affective reactions, the stronger of which are potentially harmful such as panic or maybe fear, and the weaker of which might support administrative measures that restrict liberties for a period of time.

Methods

EMA Design

The aim of this study was predominately descriptive as it intended to document levels and changes in worry at the individual level during the early phase of the COVID-19 pandemic in Sweden. The analysis also pursued a secondary aim to demonstrate that data collected by a smartphone with an EMA tool in daily rhythm can be a basis for meaningful analyses of the formation and change of emotions toward a phenomenon such as the COVID-19 pandemic. EMA was developed to study mood management and has typically been applied by paper and pencil before the ubiquity of the smartphone [35-38]. Respondents are contacted by EMA, with contacts serving as a reminder that a questionnaire is due, as a device to place the time of the interview at a particular hour of the day, or in juxtaposition to particular events. A meta-analysis turned to positive perceptions of one's well-being and documented studies that asked participants questions up to 12 times a day that were to be answered on their phones [36]. Objective measures of physiological variables are compatible with the methodology, but have not been employed very often to date.

Huckins et al [38] took the chance to add another wave of interviews to an ongoing study of students' mental health to assess the reactions toward the pandemic in the spring of 2020, employing smartphone EMA technology. They found students to be more depressed and more anxious than they had been in a comparison period prior to the pandemic. Our aim was to demonstrate the suitability of smartphone-based EMA technology beyond the study of mood.

Recruitment and Sample

Initially, students at the Department of Psychology, Umeå University, Sweden, were informed of the study via email. An article describing the study in the local newspaper in Umeå also attracted participants. PRO, an organization for retired people in Umeå, was also approached. An online invitation letter was available from March 25 to May 17, 2020. A link to the website with the invitation letter was distributed across social media all over Sweden, and participants came from the whole country. These procedures clarify that we never aimed at a representative sample, and thus we do not claim representativity.

The invitation letter detailed the purpose of the study, use of the Smartphone Ecological Momentary Assessment (SEMA^3) app [39] for distributing questionnaires, the load of questionnaires to be answered (how many, how often, how long), and informed of participants' voluntary cooperation and rights; General Data Protection Regulation (GDPR) legislation was followed. Participants received an email for downloading the app. Participants first received an introduction survey covering sociodemographic variables, followed by daily surveys with thematic questions.

Owing to the different dates of entry to and exit from the study, different lengths of participation in the study, and different spacing between surveys, our sample presents unbalanced panel data [40]. Overall, 328 adults participated in the survey; given that three survey waves are considered to be the minimum for

performing a multilevel analysis [41], we dropped all participants who only completed the first introductory survey and one or two additional surveys (n=68). 31% (21/68) completed only one survey, 31.5% (22/68) completed two, and 37% (25/68) completed three surveys, including the introduction survey. For inclusion in the study, participants had to (1) be of adult age (18 years or above), (2) be fluent in Swedish, and (3) have access to a smartphone (or a tablet).

The mean age of the analytical sample was 49.6 (SD 15.75) years, 76.7% (201/262) were women, and 80.9% (212/262) had received an academic education or were enrolled in such an institution at the time of the survey.

No gender, age, and education differences were found between dropouts and remaining participants, the latter of whom were included in the final sample (N=262). Participants in the final sample handed in between 4 and 28 surveys, including the introduction survey. This amounted to a mean of 12.3 (SD 5.46) surveys per participant. Table 1 presents an overview of how many participants replied to how many surveys. The total number of surveys we received from the sample of respondents was 3226. The analyses reported below are based on these cases.

Table 1. Number of questionnaires filled in per participant (N=268).

Number of completed interviews	Participants, n (%)
4	18 (6.7)
5	14 (5.2)
6	11 (4.1)
7	21 (7.8)
8	12 (4.5)
9	23 (8.6)
10	10 (3.7)
11	10 (3.7)
12	10 (3.7)
13	32 (11.9)
14	10 (3.7)
15	18 (6.7)
16	17 (6.3)
17	18 (6.7)
18	8 (3.0)
19	11 (5.2)
20	4 (1.5)
21	10 (3.7)
22	3 (1.1)
23	5 (1.9)
24	1 (0.4)
25	1 (0.4)
26	0 (0)
27	0 (0)
28	1 (0.4)

Data Collection

We used the SEMA^3 tool [39], a readily downloadable app at no charge to participants who possessed a smartphone with either the Android or iOS operating system. The tool was developed by a group of researchers at the Melbourne School of Psychological Sciences and is suitable for conducting intensive longitudinal survey research [42]. This tool allows for delivering surveys at fixed points in time or at fixed time intervals. During the period of data collection of almost 8 weeks, new questions were added to closely monitor the development, ask questions of relevance for participants, and assess reactions close in time to experiences, as in line with the EMA methodology.

Questions related to our dependent variable (general worry of COVID-19) and major independent variables (severity, susceptibility, and efficacy of safeguard measures) were asked every day. Other questions were scheduled at different intervals, covering aspects such as propensity for behavior change,

personal response efficacy, and social factors such as loneliness due to the pandemic. Each day at 10 AM, a survey was released and participants then had 12 hours to complete it. The first thematic survey could appear on the same day as the introduction survey. Not all participants responded on a daily basis.

After approximately 14 days, participants were thanked, irrespective of the number of surveys they had handed in. They were also asked about their experience of taking part in the study and invited to continue at a lower rate, twice a week. Nevertheless, the dependent and major independent variables were contained in every questionnaire. Assessment of governmental actions appeared in the survey for the first time on April 1, 2020. Questions on specific aspects of worry were added to the survey from April 16, 2020, onward.

No personal data were collected since each participant was assigned a code without any link to the participant's ID number or postal address. Moreover, no sensitive information was collected. A risk and vulnerability analysis was carried out in collaboration with the Information Technology Service Department at Umeå University according to a standardized protocol documenting information types and assessment of the information based on security aspects of confidentiality, accuracy, and accessibility. Thereafter, a risk analysis was performed.

Measures

Survey Design

Most survey questions were adapted from previously used measures in a small pilot study, translated from English to Swedish, and tailored to fit the study setting. Questions on specific worries and trust in government were originally formulated in Swedish. Unless otherwise noted, the following measures were assessed on a 7-point Likert scale ranging from 1 ("do not agree") to 7 ("agree").

Worry

Worry was measured by a single item question, "To what extent are you worried about the coronavirus?" on a scale from 0-10, where 0 corresponds to "not worried at all" and 10 corresponds to "very worried" (mean 6.67, SD 2.54). This question was part of every survey, including the introduction survey. Using a single item is justified by the need to keep the daily questionnaire as short as possible to reduce the number of possible dropouts.

Specific worries were assessed with respect to five items once a week as of April 16: "Are you worried about getting infected by the coronavirus yourself?" "Are you worried about someone close to you being infected by the coronavirus?" "Are you worried that your personal finances have or will be affected by the spread of the coronavirus?" "Are you worried that the Swedish economy has or will be affected by the spread of the coronavirus?" and "Are you worried that the world economy has or will be affected by the spread of the coronavirus?" (specific worries combined: mean 3.98, SD 1.84 on a scale of 1-7; α=.71). The general worry item and the combined five specific worries showed good internal consistency (Cronbach α=.76). This indicates that the single-item measure produced results of quality comparable to the scaled 5-item measure and supports the validity of the single-item measure.

Perceived Severity of COVID-19

Severity of the disease was measured by three direct questions: "Corona is a threat to everyone"; "Fighting the coronavirus is not a matter of illness or health, it is a matter of life and death"; and "There is no greater health threat than Corona right now." All three items were averaged to create a compound score (mean 5.29, SD 1.55; α=.84).

Perceived Susceptibility to COVID-19

Susceptibility was measured with three scaled questions about the risk of catching a disease: "Compared to others in my age group, I am less likely to be infected"; "I don't think my family will get infected"; and "Even when the coronavirus gets closer, I don't think I'll get it." Agreement with the statements was originally coded high, but as agreement signals low susceptibility, the three questions were reversed, with 1=low susceptibility and 7=high susceptibility. All three items were averaged to create a compound score (mean 2.51, SD 1.44; α=.84).

Efficacy of Safeguard Measures

The three items related to safeguard measures, "The actions taken so far can slow the spread of the coronavirus," "The recommendations that apply to everyday behavior will work and will reduce the spread of the coronavirus," and "Politicians responsible for public health will be able to control the spread of the coronavirus," were combined, and a compound score was created (mean 2.51, SD 1.44; α=.91).

Assessment of Government

A single-item question was used to measure how people assessed governmental performance in management of the COVID-19 pandemic: "How do you assess the government's way of handling COVID-19?" This was measured on a 10-point scale, ranging from 1 "The government makes the right decisions" to 10 "The government makes the wrong decisions" (mean 4.16, SD 2.97). This question appeared in the main survey as well as in the follow-up survey.

Data Analysis

Analysis of the longitudinal data of being worried and other related factors was performed within a multilevel modeling approach [41]. The multilevel model for change allows investigating both change within and between individuals. The analysis of the within-person change (Level 1) concerns the individual development that each subject experiences over time and is attributable to a personal combination of different influence factors, whereas the change between subjects (Level 2) is related to influence factors that are common to groups of subjects in a given sample. In this study, we were particularly interested in the question of how the changes in worry differ between subjects according to different levels of perceived severity, perceived susceptibility, efficacy of safeguard measures, and the assessment of government. The two-level hierarchical models were estimated using a maximum-likelihood method in SPSS Statistics 25. The first step of the analysis

involved the estimate of an unconditional mean model, which was followed by the unconditional growth model and other models in turn, each adding a new predictor. This procedure enables determining the relative contribution of each new variable on top of the factors that were already considered in the earlier models. More specifically, after having tested the unconditional mean (Model A) and unconditional growth (Model B) models, we added perceived severity (Model C), perceived susceptibility (Model D), efficacy of safeguarding (Model E), and finally agreement with government (Model F) to the model. All of these variables were tested as possible influence factors of both initial status and change: the intercept represents each subject's average level of worry, while the coefficient on time indicates the increase based on each additional wave the participant took part in.

Results

Descriptive statistics indicated that, on average, adults in our Swedish sample showed a level of worry of 6.20 at the very beginning of their participation in the study, which increased over time by an estimated 0.07 per day. Of note, large standard deviations were associated with both mean values, indicating that people differ widely with respect to their initial status of worry as well as with respect to their rate of change. The negative correlation coefficient between initial status and rate of change suggests that those with higher levels of worry at the beginning increased their worry level less rapidly compared with those who were initially less worried.

Table 2 presents an overview of all models that were tested. Model A represents the unconditional mean model, which provides information about the variation of the outcome worry. This model does not include either a time variable or any predictor. The mean value of worry across all occasions and individuals was 6.54 (on a scale from 0 to 10), indicating that the study participants were worried to some extent, between the two extreme values. The estimated within-person variance was 0.77, indicating that people do change their level of worry to some extent; between-person variance was 5.38, indicating a large amount of variation in worry among the individual participants of the study. The intercept as well as the two variance components were significant at the P<.001 level, meaning that adding additional variables may reduce the magnitude of the two variance components.

Model B, also called the unconditional growth model [41], includes the participation time of the individuals to the study as an additional predictor, enabling quantifying differences between participants with respect to the rate of change in their worries. According to Model B, the average change trajectory of participants had an intercept of 6.29 (P<.001) in worry and a slope of .045, significant at the P<.001 level, indicating that the level of worry increased between the end of March and the beginning of May. The within-person variance component (0.56, P<.001) of Model B summarizes the extent to which the data vary around the individual linear change trajectory (not around the person-specific mean), whereas the two variance components on level 2, initial status and rate of change, estimate the between-person variability in initial status (5.25, P<.001) and rates of change (0.015, P<.001). Adding other factors that would reduce the amount of variability in these components would help to improve the model fits. The fact that Model B is a better fit than Model A can be derived from a direct comparison, as shown by the values for R^2_e and R^2_0. The former represents the within-person residual in model A and B: comparing both values showed a decline of 27% (0.77–0.56/0.77=0.27), meaning that 27% of the variance is explained by introducing the time variable. The second value, R^2_0, represents the variance component at the outset. Comparing both models indicates an improvement of Model B with respect to this component by 2% (5.38–5.25/5.38=0.02). Therefore, including a time variable in the model particularly improved the estimate of worries at the outset.

The covariance component quantifies the association between the initial status of worry and its development over time. As such, this component helps to answer the question of whether people who are more worried at the beginning also become more (or less) worried over time. Reexpressing the covariance as a correlation coefficient [41], the relationship amounts to –0.16, meaning that those who were worrying more at the beginning are becoming slightly less worried over time. Overall, Model B showed that some of the within-person variation is associated with time, and that most of the variability in worry resided between the participants at the start, with only a small amount of variability, albeit significant, found in the change over time.

Model C and Model D bring in the aspects of threat appraisal. To facilitate interpretation, we centered perceived severity on its sample mean (5.044); to avoid giving individuals who participated in more waves greater weight in the model, mean centering was performed on the person-level data. Therefore, both intercepts, that of initial status (6.43) as well as that of the rate of change (0.03), represent the average fitted values, which were both significant at the P<.01 level. Participants with an average value of perceived severity showed a value of initial status that was 0.03 points higher. The estimated rate of change in worry for a participant with an average level of perceived severity of COVID-19 amounted to 0.01. Although this is a fairly small level of increase, it was significant at the P<.01 level, suggesting that during the study period, participants, on average, increased their level of worries about COVID-19.

Table 2. Models predicting general worry.

Parameter	Model A	Model B	Model C	Model D	Model E	Model F
Fixed effects						
Initial status						
Intercept	6.54*** (0.14)[a]	6.29*** (0.14)	6.43*** (0.14)	6.44*** (0.14)	6.43*** (0.13)	5.96*** (0.25)
Severity	N/A[b]	N/A	.20*** (.04)	.21*** (.04)	.21*** (.04)	.19*** (.04)
Susceptibility	N/A	N/A	N/A	.14*** (.04)	.13** (.04)	.14*** (.04)
Efficacy of safeguarding	N/A	N/A	N/A	N/A	–.15** (.04)	–.11*** (.03)
Agreement with government	N/A	N/A	N/A	N/A		.13** (.05)
Rate of change						
Intercept	N/A	.045*** (.01)	.03** (.01)	.03** (.01)	.03** (.01)	.04** (.02)
Severity	N/A	N/A	.01** (.005)	.01* (.005)	.01* (.005)	.01** (.004)
Susceptibility	N/A	N/A	N/A	–.01** (.004)	–.01* (.004)	–.01** (.004)
Efficacy of safeguarding	N/A	N/A	N/A	N/A	.003 (.004)	–
Agreement with government	N/A	N/A	N/A	N/A	N/A	–.002 (.003)
Variance components						
Within-person (level 1)	0.77*** (0.02)	0.56*** (0.02)	0.52*** (0.02)	0.51*** (0.02)	0.51*** (0.02)	0.50*** (0.02)
In initial status (level 2)	5.38*** (0.48)	5.25*** (0.48)	4.52*** (0.44)	4.48*** (0.44)	4.40*** (0.43)	4.36*** (0.45)
In rate of change		0.015*** (0.002)	0.013*** (0.002)	0.013*** (0.002)	0.013*** (0.002)	0.011** (0.002)
Covariance		–0.045* (0.02)	–0.065** (0.02)	–0.06** (0.02)	–0.06** (0.02)	–0.06** (0.02)
R^2_e		0.27	0.071	0.02	0	0.01
R^2_0		0.02	0.14	0.01	0.02	0.04
R^2_1		N/A	0.13	0	0	0.15
Deviance	9441	8869	7265	7238	7218	6649
AIC[c]	9447	8881	7281	7258	7242	6675
BIC[d]	9466	8918	7328	7317	7312	6750

[a]Numbers in parentheses denote the standard error.

[b]N/A: not applicable (not included in the model).

[c]AIC: Akaike information criterion.

[d]BIC: Bayesian information criterion.

*P<.05, **P<.01, ***P<.001.

Considering the variance components of Model C, we found that the within-person variance decreased from 0.56 to 0.52, which corresponds to a small reduction by 0.7%. The reduction of variance in the initial status was more remarkable at 14%, from 5.25 to 4.52, by adding severity as a predictor that explains levels of worry at the beginning of the participants' trajectory. Given that this value is significantly different from 0 (P<.001), other factors may be added to the model to explain the existing variance in Model C. Additionally, the variance component of rate of change, R^2_1, diminished by 13% (from 0.15 to 0.13) by introducing the predictor of perceived severity. Given that R^2_1 remained significantly different from 0, other predictors may still reduce the amount of the variance in this component of Model C.

In Model D, we added susceptibility, the other component of threat appraisal, which should further explain why people increase their worries over time. As in the previous analysis, we used the mean-centered value of susceptibility (5.401). This addition brought forth the following conclusions. First, controlling for the effects of susceptibility on initial status and rate of change, the effects of severity on initial status and rates of change on participants' worries amounted to .21 (P<.001) and .01 (P<.05), respectively. Second, keeping the value of severity constant, the effects of susceptibility on initial status and rates of change on participants' worry amounted to .14 (P<.001) and –.01 (P<.01), respectively, meaning that participants who differed by one point on perceived susceptibility at the initial status showed higher levels of worry by .14. Even if they were more worried at the beginning, their

average rate of change was .01 lower, indicating that participants who believed they were more susceptible at the beginning revealed a slower rate of increase of worry over time compared with those who felt less worried at the initial status; in other words, susceptibility was negatively associated with the rate of change in worry.

When we added susceptibility as a predictor of initial status of worry as well as of the rate of change, the amount of variance also shrank to some extent. The within-person variance was reduced from 0.52 to 0.51 (P<.001), the initial state variance dropped by approximately 1% point from 4.52 to 4.48 (P<.001), and the rate of change variance remained unchanged.

To improve the model further, we added perceived e*fficacy of the safeguardmeasures.* Again, we mean-centered the variable (perceived efficacy mean 5.3) to facilitate interpretation of the coefficients. Considering that the addition of perceived efficacy of safeguarding in Model E indicates an effect on levels of worry in the expected direction, holding perceived severity and susceptibility constant, two people who differ by 1 point in their view of whether safeguards were effective or not did show a difference in the level of worry by –.15. In other words, the less people were convinced that safeguard measures were effective, the more were they worried about COVID-19. Although this effect was significant at P<.001, no effect was found with respect to the rate of change (.479), meaning that people did not change their minds about safeguards measures. In the following model, we therefore dropped the perceived efficacies of safeguard measures as a predictor of the trajectory, but not as a predictor of initial status. Given the impact of attributed efficacy of safeguard measures on the initial status of worry, the corresponding variance component in Model E shrank from 4.48 to 4.40, representing a decrease of almost 2%.

The final model, Model F, added the new predictor of to what extent people think that the government properly handled the COVID-19 crisis in Sweden. Assessment of government predicted the initial status of worry but not the change of worry over time: keeping all other variables (perceived severity, susceptibility, and efficacy of safeguarding measures) constant, lower levels of agreement with government measures (expressed by the low end of a scale of 1-10) indicated a higher level of being worried about the virus (β=.13, P=.008). Given that assessment of government performance did not change the trajectory of worry over time, we finally excluded this variable as a predictor of rate of change.

The deviance statistics, including the Bayesian information criterion and Akaike information criterion, indicate how the models improve by adding the single variables. Additionally, as recommended by Singer and Willett [41], the pseudo-R statistic was computed for the within-person variance, the initial status, as well as for the rate of change variance components to show how the variance components decreased from model to model, which indicates a growing quality of the model.

Discussion

Principal Findings

Foremost, this analysis demonstrates the suitability, and perhaps even the necessity, of this type of statistical model building. Further, according to our formulated aims with regard to EMA methodology, this study also shows that meaningful data collection can be achieved by employing EMA along with using smartphones to collect data. There are several aspects to consider in interpreting the model-building results.

Research to date tells us that the self-perceived susceptibility to fall victim to a threat, the perceived severity of a risk (ie, the damage it can do), the belief in the efficacy of institutional safeguarding measures taken, and trust in government or other institutions responsible for public health are among the causal factors of risk assessment and related variables. The first noteworthy general result of our study is that the particular case of the COVID-19 pandemic confirms previous results with a very special context, a different study design (ie, a longitudinal perspective), and a new type of data analysis as a crucial methodological innovation. The predictors mentioned have a strong cognitive component. In contrast, worries and their development over time belong to the factors with a strong affective component. How the two are linked is not yet clear and needs to be addressed in future research.

This innovation carries potential to enlarge the analytical perspective. When we also look at the data taking temporal development into account, we found an impact of more or less the same variables as those that emerged in previous cross-sectional studies. The impact was actually found for all causal variables when the distribution of the worries at the beginning of the study period was predicted. This similarity is the second general result.

However, it was not only the worries in the beginning alone that were affected but there were also effects on the trends in the development of worries. These effects represent the third important general result, which clearly suggests that a cross-sectional analytic design would have missed an important part of the reality of people’s thinking about COVID-19 and the dynamic nature of the predictors. The fact that not all predictors produced linear trends shows that differentiation is called for, which we highlight as the fourth main finding. In our case, the contribution of perceived efficacy of the safeguarding measures and the support for the government did not have an impact on the trajectory of worry, and the impact on the initial status of worry must be considered to be modest. This might be surprising due to the rapid increase in COVID-19 mortality rates in Sweden and the intense debate regarding the Swedish policy. However, this might be linked to the largely unchanged policy in Sweden during the period under study, as well as a generally high and stable trust among Swedes in the public health agency, the health care system, other societal authorities and institutions as well as in the government [43]. Another reason for the small effects on worry trends is that very few participants stayed on board and handed in information over the entire study period of approximately 8 weeks. The average number of surveys handed in was 12.3. Therefore, even

if significant effects were shown, the possibility to demonstrate larger shifts in the trend was small.

As stated above, effects on initial status appear to be stronger and more stable than effects on rate of change. The effect on initial status is the totality of everything that happened between the start of consciousness about the pandemic and a person's entry into the study. Mass communication effects have long been known to be particularly strong in the early phases of an issue. The earliest phase was not included in our analysis. However, the influences in this pre-early phase may have overshadowed later effects.

Worries are clearly affected by perceptions of the threat, which is determined by scientific observations. Mustering the defensive forces to decide how to protect people from a health threat might leave more worries than the situation demands. The finding that agreement with government emerged as the weakest predictor might also have to do with the high trust in authorities and government in Sweden, especially during the period of data collection in this study [43].

Among the interesting observations from this study was the high level of spread in the worry measure. This can, as a matter of speculation, be considered a consequence of the participants being introduced into the study continuously almost over the complete data collection period. Those recruited early replied in a situation when there were still very few COVID-19 cases in the country. In addition, the question about assessment of government had a different point of reference at different times as some regulations were altered. The comparatively small change over time might have been due to a rather large share of participants taking part for quite a short period of time only. However, the daily questions might have mitigated the level of worry. We cannot exclude some automatic response behavior among participants in our study. Taking time for reflecting on our questions during a limited time each day, and in a format that was easy to handle, might have decreased the participants' worry for the rest of the day.

The classic factors of protection motivation, severity, and susceptibility considered in our study may have been overshadowed by other variables such as demographic and attitudinal variables [25] and response efficacy (ie, the belief in the benefits of one's own and the government's measures) [44,45].

The long Swedish tradition of political stability and trust in government may have contributed, along with a broad consensus among the political parties, at least early in the pandemic, to an impression that the fight against the virus was in good hands.

Limitations

A study with so many measurement points in time has to rely on modern digital technology, which makes it somewhat of a challenge to control sampling. For practical purposes, it might be best not even to try to control access to the questionnaire and to filing the responses. This means that a sample for such a study can take the form of a panel, with repeated application of the questionnaire to the same people; however, this will result in erratic schedules of participation and high differences in the number of days a person filed answers.

Since three interviews are generally considered the minimum for a multilevel analysis, and these three were collected within 2 days in some cases, there was not much time available to witness change. In fact, this is not a disadvantage as the pace of change is an open research issue, which could be advanced with data such as those collected in this study at times of a pandemic threat, given that individuals report during longer time periods. Moreover, it is a limitation that our analytical sample cannot be claimed to be representative of the source population. It is also a limitation that we could not ask about residency for anonymity reasons. This means we cannot link the responses on perceptions to a local or regional condition, in particular the number of people infected. This should be considered when interpreting the results.

Other limitations were technical problems with the SEMA app, such as difficulties in downloading the app and that participants did not receive any survey in some instances, or the response disappeared if the participant received a phone call at the same time. The result was that many were interested but could not participate and that some information might be lost. We had no information about participants' COVID-19 infection status, and therefore no conclusions can be drawn about the impact of an infection on the level of worry.

Limitations also originate from data that do not quite meet the standards to be applied to survey research when used to produce a final word on a controversial issue, in which all insecurity in the meaning of a study had better be avoided. Representative samples and validated measures produce results that are easier to trust than those obtained with our study design. Nevertheless, our research question does not live up to standards suitable for more final research.

Conflicts of Interest

None declared.

References

1. Vaughan E, Tinker T. Effective health risk communication about pandemic influenza for vulnerable populations. Am J Public Health 2009 Oct;99(Suppl 2):S324-S332. [doi: 10.2105/AJPH.2009.162537] [Medline: 19797744]
2. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the influenza A (H1N1) pandemic in the Netherlands: results of

three consecutive online surveys. BMC Public Health 2011 Jan 03;11:2 [FREE Full text] [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]
3. Graffigna G, Barello S, Savarese M, Palamenghi L, Castellini G, Bonanomi A, et al. Measuring Italian citizens' engagement in the first wave of the COVID-19 pandemic containment measures: A cross-sectional study. PLoS One 2020;15(9):e0238613 [FREE Full text] [doi: 10.1371/journal.pone.0238613] [Medline: 32915822]
4. Rogers R. Cognitive physiological processes in fear appeals attitude change: a revised theory of protection motivation in , Petty, RE. editors. Social Psychophysiology: a sourcebook. Guilford Press; 1983. 153-176. In: Cacioppo JT, Petty RE, editors. Social psychology: a sourcebook. New York: Guilford Press; 1983:153-176.
5. Rogers RW. A protection motivation theory of fear appeals and attitude change 1. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
6. Maddux JE, Rogers RW. Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. J Exp Soc Psychol 1983 Sep;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9]
7. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Commun Monogr 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
8. Witte K. Fear control and danger control: A test of the extended parallel process model (EPPM). Commun Monogr 2009 Jun 02;61(2):113-134. [doi: 10.1080/03637759409376328]
9. Witte K. Fear as motivator, fear as inhibitor: using the extended parallel processing model to explain fear appeal successes and failures. In: Anderson PA, Guerrero LK, editors. Handbook of communication and emotion: research, theory, application, and contexts. New York: Academic Press; 1998:423-450.
10. Rosenstock IM. Historical origins of the Health Belief Model. Health Educ Behav 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
11. Harper CA, Satchell LP, Fido D, Latzman RD. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
12. McKay D, Yang H, Elhai J, Asmundson GJ. Anxiety regarding contracting COVID-19 related to interoceptive anxiety sensations: The moderating role of disgust propensity and sensitivity. J Anxiety Disord 2020 Jun;73:102233 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102233] [Medline: 32442880]
13. Presti G, Mchugh L, Gloster A, Karekla M, Hayes S. The dynamics of fear at the time of covid-19: a contextual behavioral science perspective. Clin Neuropsych 2020;17(2):65-71. [doi: 10.36131/CN20200206]
14. Shah K, Bedi S, Onyeaka H, Singh R, Chaudhari G. The role of psychological first aid to support public mental health in the COVID-19 pandemic. Cureus 2020 Jun 25;12(6):e8821-e8829 [FREE Full text] [doi: 10.7759/cureus.8821] [Medline: 32742836]
15. Zolotov Y, Reznik A, Bender S, Isralowitz R. COVID-19 fear, mental health, and substance use among Israeli university students. Int J Ment Health Addict 2020 Jun 15:1-7. [doi: 10.1007/s11469-020-00351-8] [Medline: 32837432]
16. Asmundson GJG, Taylor S. Coronaphobia: Fear and the 2019-nCoV outbreak. J Anxiety Disord 2020 Mar;70:102196 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102196] [Medline: 32078967]
17. Asmundson GJG, Taylor S. How health anxiety influences responses to viral outbreaks like COVID-19: What all decision-makers, health authorities, and health care professionals need to know. J Anxiety Disord 2020 Apr;71:102211 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102211] [Medline: 32179380]
18. González-Sanguino C, Ausín B, Castellanos M, Saiz J, López-Gómez A, Ugidos C, et al. Mental health consequences during the initial stage of the 2020 Coronavirus pandemic (COVID-19) in Spain. Brain Behav Immun 2020 Jul;87:172-176 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.040] [Medline: 32405150]
19. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
20. Jungmann SM, Witthöft M. Health anxiety, cyberchondria, and coping in the current COVID-19 pandemic: Which factors are related to coronavirus anxiety? J Anxiety Disord 2020 Jun;73:102239 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102239] [Medline: 32502806]
21. Taylor S. The psychology of pandemics. London, UK: Cambridge Scholars; 2019.
22. Zettler I, Schild C, Lileholt L, Böhm R. Individual differences in accepting personal restrictions to fight the COVID-19 pandemic: Results from a Danish adult sample [Preprint]. PsyArXiv. 2020. URL: https://psyarxiv.com/pkm2a/ [accessed 2021-11-20]
23. Seçer İ, Ulaş S. An investigation of the effect of COVID-19 on OCD in youth in the context of emotional reactivity, experiential avoidance, depression and anxiety. Int J Ment Health Addict 2020 Jun 13:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00322-z] [Medline: 32837429]
24. Yeung DY, Fung HH. Age differences in coping and emotional responses toward SARS: a longitudinal study of Hong Kong Chinese. Aging Ment Health 2007 Sep;11(5):579-587. [doi: 10.1080/13607860601086355] [Medline: 17882596]
25. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]

26. Mucci F, Diolaiuti F, Mucci N. Lockdown and isolation: psychological aspects of COVID-19 pandemic in the general population. Clin Neuropsychiatry 2020;17(2):63-64. [doi: 10.36131/CN20200205]
27. Jungmann SM, Witthöft M. Health anxiety, cyberchondria, and coping in the current COVID-19 pandemic: Which factors are related to coronavirus anxiety? J Anxiety Disord 2020 Jun;73:102239 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102239] [Medline: 32502806]
28. Li HY, Cao H, Leung DYP, Mak YW. The psychological impacts of a COVID-19 outbreak on college students in China: a longitudinal study. Int J Environ Res Public Health 2020 Jun 02;17(11):3933 [FREE Full text] [doi: 10.3390/ijerph17113933] [Medline: 32498267]
29. 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]
30. Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]
31. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
32. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
33. Giritli Nygren K, Olofsson A. Managing the Covid-19 pandemic through individual responsibility: the consequences of a world risk society and enhanced ethopolitics. J Risk Res 2020 Apr 23;23(7-8):1031-1035. [doi: 10.1080/13669877.2020.1756382]
34. WHO official: Sweden's policy of individual responsibility "a model" for the rest of world. Foundation for Economic Education. 2020 Apr 30. URL: https://fee.org/articles/who-official-sweden-s-policy-of-individual-responsibility-a-model-for-the-rest-of-world/ [accessed 2020-01-03]
35. Haedt-Matt AA, Keel PK. Revisiting the affect regulation model of binge eating: a meta-analysis of studies using ecological momentary assessment. Psychol Bull 2011 Jul;137(4):660-681 [FREE Full text] [doi: 10.1037/a0023660] [Medline: 21574678]
36. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021 Oct 23;22(5):2361-2408 [FREE Full text] [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
37. Romanzini CLP, Romanzini M, Batista MB, Barbosa CCL, Shigaki GB, Dunton G, et al. Methodology used in ecological momentary assessment studies about sedentary behavior in children, adolescents, and adults: systematic review using the checklist for reporting ecological momentary assessment studies. J Med Internet Res 2019 May 15;21(5):e11967 [FREE Full text] [doi: 10.2196/11967] [Medline: 31094349]
38. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
39. Koval P, Hinton J, Dozo N, Gleeson J, Alvarez M, Harrison A, et al. SEMA3: Smartphone Ecological Momentary Assessment, Version 3. URL: https://sema3.com/ [accessed 2021-11-20]
40. Biørn E. Econometrics of panel data. Oxford: Oxford University Press; 2017.
41. Singer JD, Willett JB. Applied longitudinal data analysis: modelling change and event occurrence. Oxford, UK: Oxford University Press; 2003.
42. Research Tools: SEMA. The University of Melbourne. URL: https://psychologicalsciences.unimelb.edu.au/research/research-initiatives/our-work/feel-research-lab/resources [accessed 2021-11-08]
43. Public trust, thoughts and behavior during the corona crisis [in Swedish]. Kantar Sifo. URL: https://www.kantarsifo.se/sites/default/files/reports/documents/allmanhetens_tillit_tankar_och_beteende_under_coronakrisen_8_maj_2020_kantarsifo.pdf [accessed 2021-07-10]
44. Floyd D, Prentice-dunn S, Rogers RW. A meta-analysis of research on protection motivation theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
45. Milne S, Sheeran P, Orbell S. Prediction and intervention in health-related behavior: a meta-analytic review of protection motivation theory. J Appl Social Pyschol 2000 Jan;30(1):106-143. [doi: 10.1111/j.1559-1816.2000.tb02308.x]

Abbreviations

EMA: ecological momentary assessment
FCV-192: Fear of COVID-19 Scale
GDPR: General Data Protection Regulation

SARS: severe acute respiratory syndrome
SEMA: Smartphone Ecological Momentary Assessment app

Edited by R Kukafka; submitted 23.12.20; peer-reviewed by M Katapodi, D Walker, H Hochheiser, L Shen; comments to author 15.03.21; revised version received 06.04.21; accepted 14.06.21; published 29.11.21.

Please cite as:
Schulz PJ, Andersson EM, Bizzotto N, Norberg M
Using Ecological Momentary Assessment to Study the Development of COVID-19 Worries in Sweden: Longitudinal Study
J Med Internet Res 2021;23(11):e26743
URL: https://www.jmir.org/2021/11/e26743
doi:10.2196/26743
PMID:34847065

Original Paper

Examination of a Canada-Wide Collaboration Platform for Order Sets: Retrospective Analysis

Arshia Pedram Javidan[1,2], MD, MSc; Allan Brand[3], MSc; Andrew Cameron[4], LLM, MD; Tommaso D'Ovidio[3], MSc; Martin Persaud[3], CA, CPA; Kirsten Lewis[3], BSCN, MSN; Chris O'Connor[5], MD

[1]Division of Vascular Surgery, Department of Surgery, University of Toronto, Toronto, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Think Research, Toronto, ON, Canada

[4]Department of Emergency Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[5]Trillium Health Partners, Mississauga Site, Mississauga, ON, Canada

Corresponding Author:
Arshia Pedram Javidan, MD, MSc
Division of Vascular Surgery
Department of Surgery
University of Toronto
Toronto General Hospital
200 Elizabeth St
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 3131
Email: arshia.javidan@mail.utoronto.ca

Abstract

Background: Knowledge translation and dissemination are some of the main challenges that affect evidence-based medicine. Web 2.0 platforms promote the sharing and collaborative development of content. Executable knowledge tools, such as order sets, are a knowledge translation tool whose localization is critical to its effectiveness but a challenge for organizations to develop independently.

Objective: This paper describes a Web 2.0 resource, referred to as the collaborative network (TCN), for order set development designed to share executable knowledge (order sets). This paper also analyzes the scope of its use, describes its use through network analysis, and examines the provision and use of order sets in the platform by organizational size.

Methods: Data were collected from Think Research's TxConnect platform. We measured interorganization sharing across Canadian hospitals using descriptive statistics. A weighted chi-square analysis was used to evaluate institutional size to share volumes based on institution size, with post hoc Cramer V score to measure the strength of association.

Results: TCN consisted of 12,495 order sets across 683 diagnoses or processes. Between January 2010 and March 2015, a total of 131 health care organizations representing 360 hospitals in Canada downloaded order sets 105,496 times. Order sets related to acute coronary syndrome, analgesia, and venous thromboembolism were most commonly shared. COVID-19 order sets were among the most actively shared, adjusting for order set lifetime. A weighted chi-square analysis showed nonrandom downloading behavior (P<.001), with medium-sized institutions downloading content from larger institutions acting as the most significant driver of this variance (chi-gram=124.70).

Conclusions: In this paper, we have described and analyzed a Web 2.0 platform for the sharing of order set content with significant network activity. The robust use of TCN to access customized order sets reflects its value as a resource for health care organizations when they develop or update their own order sets.

(J Med Internet Res 2021;23(11):e26123) doi:10.2196/26123

KEYWORDS

evidence-based medicine; health informatics; knowledge translation; order sets; Web 2.0

Introduction

Background

There continues to be increased application of evidence-based medicine in clinical practice [1]. However, the volume of published evidence available makes direct application of the most appropriate evidence difficult to reliably apply at the point of care. Effective knowledge translation could address this problem by ensuring that physician practice reflects the best current evidence [2-4]. Knowledge transfer, a key component of knowledge translation, focuses on how information, knowledge, and resources are disseminated and exchanged among relevant clinicians. Continuing medical education (CME) is a mechanism that facilitates knowledge transfer and, more broadly, knowledge translation [4]. Despite its intuitive appeal, studies on didactic CME activities (eg, grand rounds) do not show significant changes in physician behavior [5]. Interactive CME activities can be more effective; however, their impact is limited because of narrow outreach, logistics, and cost [5-7].

Web 2.0 platforms (ie, web-based platforms that facilitate information sharing through user-generated content) allow for improved physician collaboration and knowledge translation [8,9]. For example, the Twitter Free Open Access Medical Education community generated >1 billion tweet impressions among nearly 50,000 users over 2 years [10-12]. However, these platforms predominantly focus on referential knowledge (ie, information a physician refers to through textbooks or articles) instead of executable knowledge (ie, information converted into tools used directly in patient care). The literature evaluating collaboration on these platforms has focused on platforms that primarily share referential knowledge, general use platforms (eg, Twitter), or platforms for specific specialties (eg, emergency medicine blogs) [10-15]. There have been limited studies on platforms that focus on sharing executable knowledge [9].

Order sets (collections of architected predefined orders) are a type of executable knowledge designed to deliver evidence-based best practices [16] that have been shown to improve patient care, safety, and efficiency [17-23]. Order sets are predetermined templates that represent a collection of orders specific to a particular hospital process (eg, admission to the intensive care unit) or a particular condition (eg, order set for acute coronary syndrome and order set for insulin administration in the context of diabetes or hyperglycemia). They can be either paper or electronic based and often represent best practices for the condition or process to which they pertain [17-23]. They offer benefits over traditional CME by making best practices directly actionable at the point of care in a structured format. Order sets must often be localized to meet local resource and workflow needs. This localization has typically been done on an isolated basis with no formal collaborative infrastructure among organizations. This siloed approach impedes effective knowledge transfer and, by association, knowledge translation.

Objectives

The aim of this study is to first describe a Web 2.0 network (the collaborative network [TCN]) that enables the sharing of localized executable knowledge through order sets and clinical guidelines. The variation created through localization acts as a network attractor, as clinicians may be interested in understanding how others have translated evidence into practice. We also aim to examine the use of TCN through network analysis and the use and provision of order sets in TCN stratified by organizational size. To our knowledge, this is the first study of its kind to describe the networked sharing of executable content at this scale, focusing on a Canadian setting.

Methods

The Collaborative Network

TCN acts as a resource repository for all content developed by Think Research and the organizations participating in the network. In this study, we focused on the knowledge base for 3 related sets of data: reference clinical order set content developed by Think Research and adapted order set content from 2 sources—partner agencies (eg, governmental and clinical specialist groups) and local participating hospitals or health care institutions.

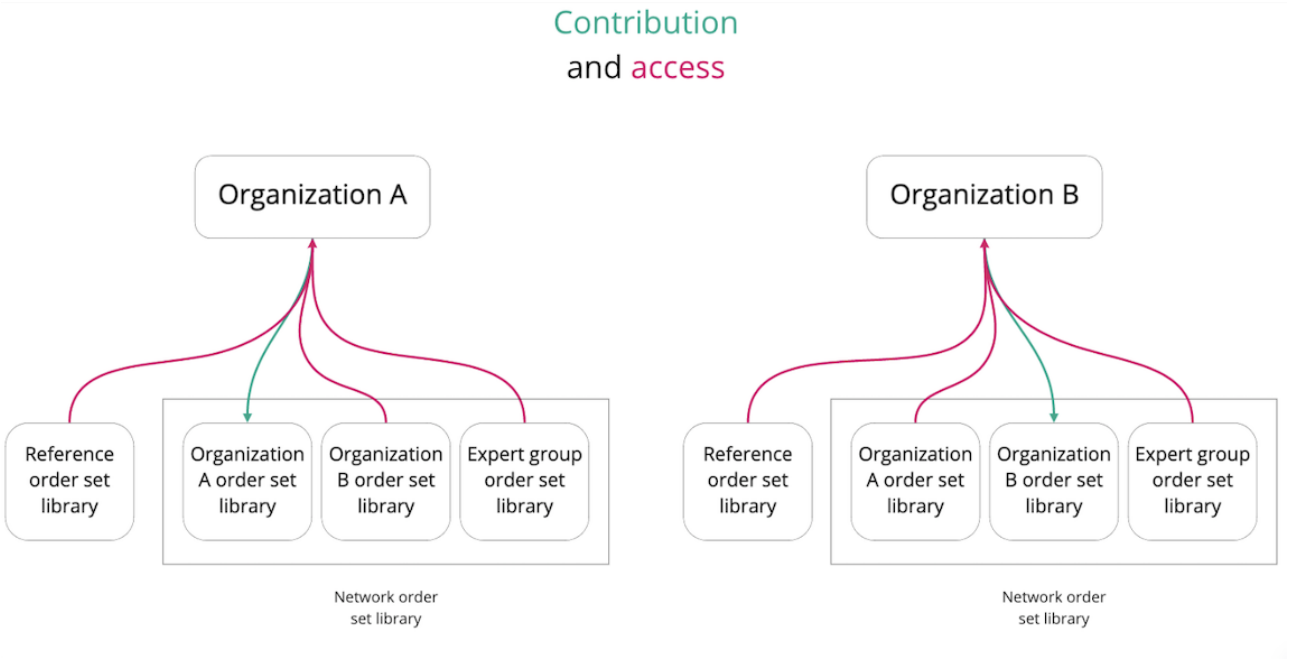
Collaboration: Contribution

TCN functions as the main system for order set knowledge translation, exchange, and collaboration across and within organizations on the network. Upon developing new order sets or updating existing order sets, participating organizations can access the network order set library for any other organization. Knowledge from one organization about how orders are localized to practice can then be accessed by other organizations. Some organizations may choose to use these order sets as the starting point for development or simply reference them during the process of updating their existing order sets. The process of contribution and access creates a community of practice in which hospitals exchange ideas about the uptake of evidence-based best practices and how it is implemented.

Collaboration: Access

Users of TCN log in to the application and search for order sets through the reference order set library, network order set library, or their organization's order set library (Figure 1). The reference order set library contains order sets that are developed by the Think Research Clinical Research and Development team, which consists of physicians, nurse practitioners, nurses, pharmacists, and other clinicians. The network order set library contains all of the order sets that are developed by partner agencies (eg, governmental and clinical specialist groups) and local participating hospitals or health care institutions. Depending on the type of order set the user is interested in, the user would select the relevant library, and results can be refined by filtering with diagnosis (eg, community-acquired pneumonia), hospital location (eg, emergency department), and other keywords. The user also has the option to use the search field (queries are made by title and keyword).

Figure 1. Contribution and access activity on the collaborative network, simplified using a model with only 2 organizations.



Ethics Approval

As this study was considered a program evaluation, it did not fall under the auspices of the research ethics board and was exempt from research ethics board review. In addition, the study was considered as using data as a secondary use that does not involve any identifier information that is specific to an individual.

Study Setting and Participants

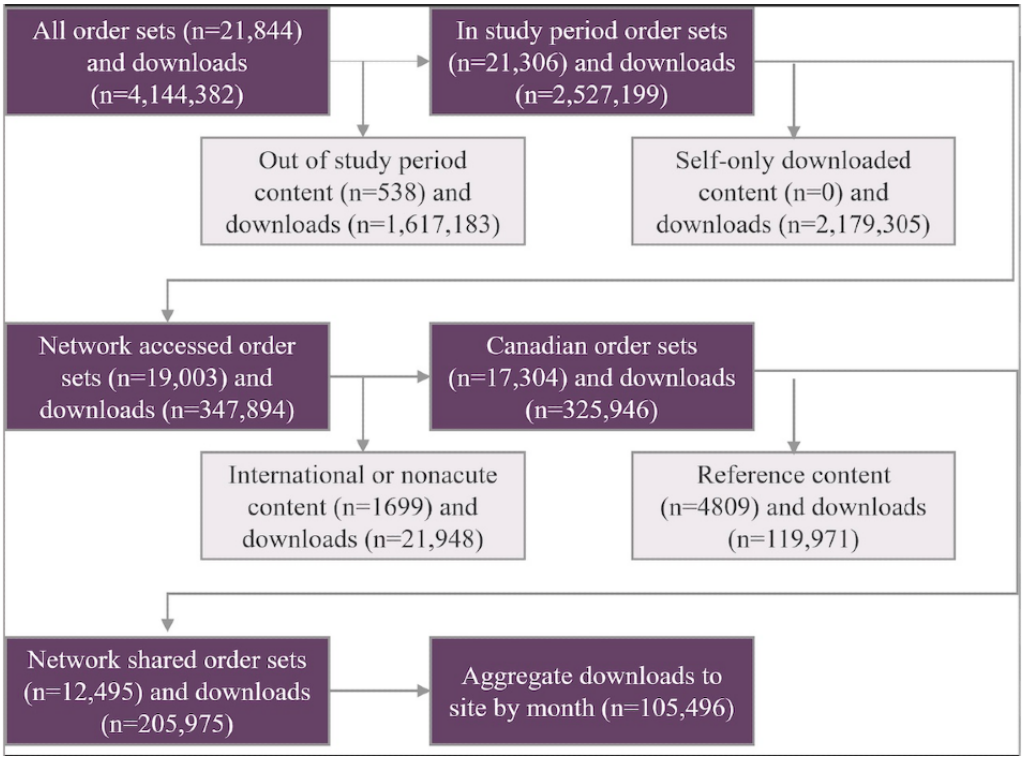
Metadata for all participating Canadian hospitals and institutions, including hospital name, location, and number of beds, were available on the centralized network. These data were collected for all participating Canadian hospitals and institutions accessing the platform from January 2015 to March 2020. Although various countries make use of the platform, we sought to describe the network in a Canadian context. Data not pertaining to Canadian users were removed from our analysis. Hospital and organization names were anonymized to preserve hospital confidentiality.

Data Collection

Filters were sequentially applied to the entirety of the data set to identify eligible order sets. These selection criteria were applied broadly across the entirety of TCN or order set library. A summary of the selection criteria for order set selection is displayed in Figure 2. As our intention was to understand interorganizational behavior, we excluded intraorganizational downloads. We also did not include those downloads from the reference library (ie, Think Research) to highlight sharing among health care organizations instead of content downloads from the reference library.

We collected metadata for all downloaded content, including order set title and document, associated content downloads (eg, clinical protocols and supporting documents), downloading hospital or institution, and owner hospital or institution. All order sets and supporting documents were coded into either a medical diagnosis or medical process, if a diagnosis was not applicable. Hospitals were categorized as small (≤70 beds), medium (> 70 beds but <300), large (>300 beds), or as *group* if the organization consisted of more than one hospital or health care institution.

Figure 2. Data selection criteria and associated number of resources for the final data set.



Descriptive Analysis

This study focused on understanding the broad patterns of network behavior across organizations. To that end, we focused on understanding trends in the types of content shared at the level of diagnosis and institution so as to avoid inferences about specific institutional trends. To reduce the impact of multiple downloads from a site over a short period by the same user, we aggregated all downloads within a month by a single user at an institution into a single action per month.

We aimed to describe overall content availability, lifetime highly shared content, and active content (defined by the number of shares over the lifetime of the content). We also examined the properties of the hospital institution's origin and downloader to describe high-level trends in sharing.

Network Analysis

We also aimed to understand if variances in institutional size were related to a site's tendency to be a source (provider of the order set) or sink (user of the order set) for content. We chose not to include each institution's health information system or geography to preserve anonymity in the analysis. We performed a weighted chi-square analysis, where the expected number of downloads was estimated by the cross-product of content available at each hospital size by the total number of downloads of content. A post hoc analysis was conducted using pairwise comparisons corrected using Bonferroni adjustment. Estimated relationship strength was calculated using Cramer V.

Descriptive statistics were computed using Excel (Microsoft Excel version 16.52). Statistical analysis and network graphs with nodes were generated using R (R version 3.6.1).

Results

Descriptive Analysis

The participants of TCN consisted of 131 unique health care organizations (including hospital groups), consisting of 360 hospitals and institutions across 8 provinces in Canada. Hospital organizations ranged in size from 11 to 3119 beds with a median number of 134 beds per hospital and a mean of 319.9 (SD 514.8) beds (Table 1). The total data set represents approximately 45.89% (41,906/91,325) of the total beds staffed in Canada [24]. Seven institutions did not share or download content from the platform and were excluded from the study.

During the study period, 12,495 institutional order sets were generated and shared by the participating institutions (Figure 2). Collectively, these order sets correspond to 658 unique medical categories and diagnoses (eg, congestive heart failure and acute kidney injury) and 25 hospital processes (eg, intensive care unit admission). The 10 most commonly shared order set diagnoses represented 21% (22,150/105,496) of all shares (Table 2). Adjusting for the length of time content was available on the network, and rare content (ie, <5 order sets per diagnosis), content relevant to COVID-19, was among the most actively shared content (Table 3).

Table 1. Characteristics of participant health care organizations (N=131).

Health care organization characteristic	Health care organizations, n (%)
Participant health care organizations	
Small (≤70 beds)	44 (33.6)
Medium (>70 but ≤300 beds)	40 (30.5)
Large (>300 beds)	25 (19.1)
Hospital group (>1 hospital[a])	22 (16.8)
Geography	
Alberta	0 (0)
British Columbia[b]	2 (1.5)
Manitoba[c]	1 (0.8)
New Brunswick	0 (0)
Newfoundland and Labrador[d]	4 (3.1)
Nova Scotia[e]	2 (1.5)
Ontario[f]	110 (84.0)
Prince Edward Island[g]	1 (0.8)
Quebec[c]	1 (0.8)
Saskatchewan[h]	10 (7.6)
Northwest Territories	0 (0)
Nunavut	0 (0)
Yukon	0 (0)

[a]251 hospitals are part of a larger hospital group.
[b]37 hospitals.
[c]1 hospital.
[d]34 hospitals.
[e]38 hospitals.
[f]138 hospitals.
[g]7 hospitals.
[h]104 hospitals.

Table 2. Most available order sets and most downloaded order sets by diagnosis.

Order set category	Order sets, n (%)
Content available	12,495 (100)
Acute coronary syndrome	359 (2.87)
Stroke or TIA[a]	284 (2.27)
Analgesia	273 (2.18)
Venous thromboembolism	218 (1.74)
Diabetes	205 (1.64)
Labor	185 (1.48)
COPD[b]	182 (1.46)
Asthma	178 (1.42)
CHF[c]	177 (1.42)
Palliative care	159 (1.27)
Other disease conditions	10,275 (82.23)
Total downloads over lifetime	105,496 (100)
Acute coronary syndrome	4122 (3.91)
Analgesia	3319 (3.15)
Venous thromboembolism	2696 (2.56)
Alcohol use, detoxification, and withdrawal	1794 (1.70)
Palliative care	2386 (2.26)
Diabetic ketoacidosis	1891 (1.79)
Stroke or TIA	832 (0.79)
Labor	1854 (1.76)
Total parenteral nutrition	1740 (1.65)
Diabetes	1516 (1.44)

[a]TIA: transient ischemic attack.

[b]COPD: chronic obstructive pulmonary disease.

[c]CHF: congestive heart failure.

Table 3. Most active content over lifetime (N=12,495 order sets).

Diagnosis group	Order sets, n (%)	Years available, mean (SD)	Absolute activity[a]
Acute coronary syndrome	359 (2.87)	2.68 (1.44)	2092.77
Stroke or TIA[b]	284 (2.27)	1.93 (1.46)	1866.64
COVID-19	22 (0.18)	0.08 (0.04)	1470.26
Analgesia	273 (2.18)	2.88 (1.43)	1400.96
COPD[c]	182 (1.46)	2.03 (1.36)	1291.47
Venous thromboembolism	218 (1.74)	2.81 (1.45)	1213.54
Alcohol use, detoxification, and withdrawal	156 (1.25)	2.70 (1.38)	1057.76
Diabetic ketoacidosis	205 (1.64)	2.58 (1.52)	1012.35
Diabetes	148 (1.18)	2.28 (1.48)	1005.66

[a]The sum of all downloads of content within that diagnosis divided by the total lifetime of all content within that diagnosis group.

[b]TIA: transient ischemic attack.

[c]COPD: chronic obstructive pulmonary disease.

Network Analysis

There were a total of 105,496 shares among institutions between January 2015 and March 2020. Institutions shared on average 98.8 distinct order sets a mean of 574.0 times (median 200; range: 0-5371). Institutions downloaded a mean of 517.8 unique order sets (median 294; range 0-5342) for a mean of 718.9 times (median 367; range: 0-7316; detailed breakdown available in Multimedia Appendix 1). Content was available for an average of 2.32 years (median 2.32 years; range 0.08-5.25 years). Figure 3 demonstrates a network for the sharing of all order sets stratified by hospital size. Figure 4 demonstrates a network for sharing of order sets related to COVID-19 as a specific example to highlight the sharing among sites.

Figure 3. Network diagram showing total collaborative network for downloads of order sets during the study period by hospital size. Node color represents hospital size (gray=small; teal=medium; purple=large; loyal blue=group). Node size is representative of the source size of the institution (ie, popularity of that site's content). Edge color and width represent sink hospital size and the relative number of downloads (ie, thicker=more unique downloads of content). Arrows point from source hospital to sink hospital.

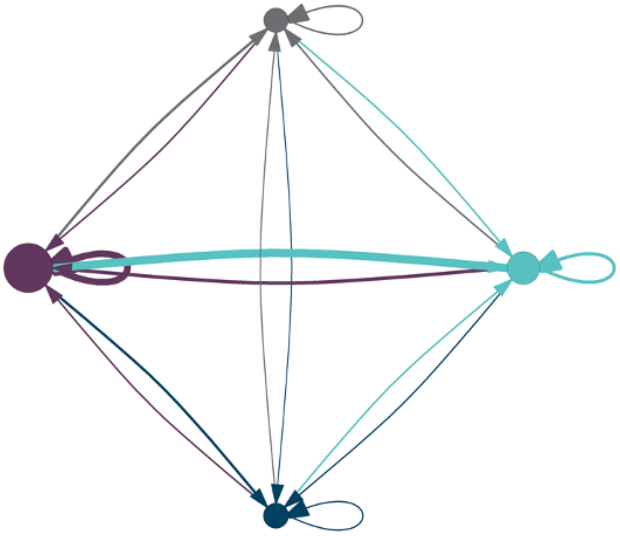


Figure 4. Network graph of order set downloads for order sets corresponding to COVID-19. Node color represents hospital size (gray=small; teal=medium; purple=large; loyal blue=group). Node size is representative of the source size of the institution (ie, popularity of that site's content for this particular order set). Edge color and width represent sink hospital size and the relative number of downloads (ie, thicker=more unique downloads of content). Arrows point from source hospital to sink hospital.

At the institution level, there was a significant (weighted χ^2_{15}=64313.9; *P*<.001) difference between the expected number of downloads by hospital size and content availability (Table 4). Medium-sized institutions downloading content from large institutions were the largest drivers of this variance (chi-gram=124.70). Post hoc analysis showed that all institution size interactions were statistically significant (Table 4), with a moderate interaction between institution size of order set providers and users (Cramer V=0.26).

Table 4. Summary of institutional sharing and content availability of 105,496 shares.

Downloader	Owner			
	Small	Medium	Large	Group
Small				
Shares, n (%)	4391 (4.16)	4046 (3.84)	6075 (5.76)	2782 (2.64)
Chi-gram	6.72[a]	–18.66[a]	–12.13[a]	–23.13[a]
Medium				
Shares, n (%)	4386 (4.16)	8518 (8.07)	21029 (19.93)	4044 (3.83)
Chi-gram	–9.98[a]	18.60[a]	124.70[a]	–19.92[a]
Large				
Shares, n (%)	4462 (4.23)	9114 (8.64)	16906 (16.03)	4632 (4.39)
Chi-gram	–32.07[a]	–7.01[a]	35.86[a]	–35.68[a]
Group				
Shares, n (%)	2074 (1.97)	3613 (3.42)	6135 (5.82)	3289 (3.12)
Chi-gram	–29.16[a]	–23.39[a]	–9.94[a]	–14.31[a]

[a]$P<.00006$ (Bonferroni adjusted $P<.001$).

Discussion

Principal Findings

We have presented a description and analysis of a Web 2.0 platform that facilitated the sharing of order sets across Canada, over a breadth of hospital sizes and clinical specialties, with >100,000 shares across 131 unique health care organizations over 5 years. This is the first study of its kind that has described the sharing of clinical decision supports, and more broadly, translation of executable knowledge at this scale with transparency. Although other studies have focused on artifacts of exchange (ie, amount of content available), this study also demonstrates the actual network of sharing supported by content [9].

Various models have been proposed to conceptualize knowledge translation. One framework developed by Graham et al [25,26] and subsequently adopted by the Canadian Institutes of Health Research defines knowledge translation as a dynamic and iterative process concerning the creation, dissemination, and application of knowledge. The foundation of this framework proposed by Graham et al [25,26] involves knowledge creation, consisting of knowledge inquiry (eg, primary research concerning a particular disease state), synthesis (eg, systematic reviews and guidelines derived from primary research), and then products or tools (eg, order sets developed by an entity such as Think Research, which is then disseminated). Knowledge transfer, the primary focus of TCN, is a fundamental step in the knowledge translation process. Central to knowledge translation is also the adoption of knowledge in the local context from a top-down perspective. This can occur with order sets as well.

Order Set Customization and Sharing

Order set customization may be done for multiple reasons, including local variations in care processes, differences in resource availability (eg, a 14-bed hospital will have different resources than a 1500-bed hospital), or ambiguity in high-level recommendations provided in clinical guidelines (eg, keeping systolic blood pressure below 160 mm Hg under certain circumstances but not specifying which drugs to use and when to start). In addition, many care processes are not always addressed in the guidelines and other medical literature. For this type of content, organizations must create their own order sets. These mechanisms all contribute to the customization of order sets and produce the variations that exist among organizations. In turn, this variation can catalyze sharing.

The most significant finding of this study is the heavy use of the network for the purpose of knowledge transfer. There were no mechanisms in place to enforce or encourage sharing of order sets. Users of the platform could opt to, for example, download a complete order set from the reference library instead of downloading from another institution. Despite this, there was extensive sharing. This may derive from a number of factors; for example, clinicians may wish to emulate the practice of a colleague, learn from the order sets of large academic centers, or inquire into how an order set was modified in a hospital with similar capacities and facilities.

Viewing another organization's order sets takes time and effort from otherwise busy health care practitioners. Health care providers would not do this if they did not derive value from viewing another organization's order sets. This downloading even occurred during a time of severe organizational stress during the early phases of the COVID-19 pandemic. Network activity also reflected these contemporary developments. For example, COVID-19 was first identified in Canada as early as January 23, 2020 [27]. However, as of March 2020, order sets related to COVID-19 have held a spot among the most actively shared order sets. This is telling of the network's capacity for rapid knowledge transfer and, more broadly, knowledge translation.

An examination of sharing patterns stratified by hospital size also revealed that although the largest content source was large hospitals, contributing to 47.4% of shares, organizations of all sizes played substantial roles in sharing. This reflects the value that all organizations, regardless of size, can contribute to a network of knowledge and sharing.

Although users of TCN had no explicit measures or incentives to do so, the sharing of order sets was the predominant mechanism of knowledge translation. This may suggest that the mere availability of the platform naturally promoted sharing. This trend of increasing uptake of Web 2.0 platforms as a means of knowledge translation among clinicians has also been seen with platforms such as Twitter and blogs [15]. As technologies continue to evolve and clinicians become increasingly interconnected, strategies for knowledge translation must adapt appropriately.

Although this study sought to document the overall trend in order set sharing across a network, it does not attempt to describe the reasons for content being accessed. There are a number of possibilities, including organizations reusing the entire order set for their own local practice, reusing components and reintegrating it into another order set, or simply using it as a general source of knowledge or ideas. In addition, although we do examine how resources were disseminated and exchanged among participants in the network, we do not examine exactly how these order sets were used in the clinical setting (eg, the proportion of downloads that led to an order set being used in the clinical setting and whether they were understood or used properly). Although knowledge transfer is a critical part of knowledge translation, implementation of knowledge in a clinical setting is also important to evaluate. Finally, there were other factors for analysis that could be evaluated, including geography, electronic health record (EHR) system, and user-specific access. Further study in these areas could produce valuable insights about the outputs as a result of the sharing of executable knowledge.

Next Steps

Looking forward, physician learning and knowledge translation will likely continue to harness the potential of Web 2.0. The response of the academic and clinical landscape to COVID-19 represents a prime example of how knowledge translation systems must adapt to the need for rapidly changing information. Platforms such as Slack and Twitter served as mechanisms to evaluate and distribute preprints of articles and processes that accelerated knowledge translation [28]. Twitter, WeChat, and university websites also served as platforms for sharing infographics and expert recommendations [29,30]. Much of the information disseminated through these means were referential in nature (eg, disease characteristics and epidemiology), with less translation of experiential knowledge or tools that could be applied for on-the-job learning [31].

Similarly, pre-existing platforms, such as information systems and EHRs, can continue to evolve to support better care plans, improve interdisciplinary communication and workflow, and provide clinical decision supports. In this process, order sets may play an increasingly important role in EHRs. The *plan-centric* EHR is forward looking and focuses on redesign to not only record patient information but also to enhance patient care and optimize the delivery of care. Seamless adoption of order sets into EHRs may be one component of this redesign; the ability to access localized order sets across geographic barriers can serve as an accelerant for the consistent integration of evidence-based guidelines in future EHR design [32].

As Web 2.0 platforms continue to evolve, ethical use must also be considered. The benefits of mass user contributions, convenience, and low barriers to entry of using Web 2.0 platforms must be considered in light of some of the possible drawbacks, such as lack of peer review and the possibility of the spread of misinformation [30,33,34]. To this end, guidelines have been developed by authors for the responsible use of social media–disseminated information [30]. In addition, these growing platforms should be viewed not necessarily as a replacement of more traditional knowledge translation activities but as a supplement. Synergistic benefits can manifest from marrying various types of platforms, producing gains for physician collaboration, CME, and enhancing experiential learning through sharing of executable knowledge.

Acknowledgments

The authors would like to thank Michael Cheung, Monica Yu, and Delphine Lim for their help in gathering and organizing the data for this analysis.

Authors' Contributions

APJ, AB, AC, TD, MP, and CO assisted with conceptualization, study design, interpreting the data, drafting the article, providing critical revisions, and approving the final version for submission. KL assisted with the interpretation of the data, drafting the paper, providing critical revisions, and approving the final version for submission. APJ, AB, TD, and MP assisted with data collection and analysis.

Conflicts of Interest

APJ has no conflicts of interest to declare. AB, AC, TD, MP, and CO were previously employed by Think Research, the source of the database used in this study. KL is currently employed by Think Research. None of the authors received any financial incentive for any of their contributions to this study. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Multimedia Appendix 1
Order set availability and access by institution.
[DOCX File , 14 KB - jmir_v23i11e26123_app1.docx]

References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. J Am Med Assoc 1992 Nov 04;268(17):2420-2425. [doi: 10.1001/jama.1992.03490170092032] [Medline: 1404801]
2. Wandersman A, Duffy J, Flaspohler P, Noonan R, Lubell K, Stillman L, et al. Bridging the gap between prevention research and practice: the interactive systems framework for dissemination and implementation. Am J Community Psychol 2008 Jun;41(3-4):171-181. [doi: 10.1007/s10464-008-9174-z] [Medline: 18302018]
3. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7:50 [FREE Full text] [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]
4. Davis D, Evans M, Jadad A, Perrier L, Rath D, Ryan D, et al. The case for knowledge translation: shortening the journey from evidence to effect. Br Med J 2003 Jul 05;327(7405):33-35 [FREE Full text] [doi: 10.1136/bmj.327.7405.33] [Medline: 12842955]
5. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? J Am Med Assoc 1999 Sep 01;282(9):867-874. [doi: 10.1001/jama.282.9.867] [Medline: 10478694]
6. Sandal S, Iannuzzi MC, Knohl SJ. Can we make grand rounds "Grand" again? J Grad Med Educ 2013 Dec;5(4):560-563 [FREE Full text] [doi: 10.4300/JGME-D-12-00355.1] [Medline: 24455001]
7. Haynes RB. Some problems in applying evidence in clinical practice. Ann N Y Acad Sci 1993 Dec 31;703:210-224. [doi: 10.1111/j.1749-6632.1993.tb26350.x] [Medline: 8192298]
8. O'Reilly T. What is Web 2.0. O'Reilly. 2005 Sep 30. URL: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html [accessed 2021-09-24]
9. Wright A, Bates DW, Middleton B, Hongsermeier T, Kashyap V, Thomas SM, et al. Creating and sharing clinical decision support content with Web 2.0: issues and examples. J Biomed Inform 2009 Apr;42(2):334-346 [FREE Full text] [doi: 10.1016/j.jbi.2008.09.003] [Medline: 18935982]
10. Roland D, Spurr J, Cabrera D. Preliminary evidence for the emergence of a health care online community of practice: using a netnographic framework for Twitter hashtag analytics. J Med Internet Res 2017 Jul 14;19(7):e252 [FREE Full text] [doi: 10.2196/jmir.7072] [Medline: 28710054]
11. Cevik AA, Aksel G, Akoglu H, Eroglu SE, Dogan NO, Altunci YA. Social media, FOAMed in medical education and knowledge sharing: local experiences with international perspective. Turk J Emerg Med 2016 Sep;16(3):112-117 [FREE Full text] [doi: 10.1016/j.tjem.2016.07.001] [Medline: 27857988]
12. Carley S, Beardsell I, May N, Crowe L, Baombe J, Grayson A, et al. Social-media-enabled learning in emergency medicine: a case study of the growth, engagement and impact of a free open access medical education blog. Postgrad Med J 2018 Feb;94(1108):92-96. [doi: 10.1136/postgradmedj-2017-135104] [Medline: 29054933]
13. Colbert GB, Topf J, Jhaveri KD, Oates T, Rheault MN, Shah S, et al. The social media revolution in nephrology education. Kidney Int Rep 2018 May;3(3):519-529 [FREE Full text] [doi: 10.1016/j.ekir.2018.02.003] [Medline: 29854960]
14. Syed S, Rosenberg H. FOAMed and social media - innovation or disruption? Can J Emerg Med 2018 Nov;20(6):818-820. [doi: 10.1017/cem.2018.471] [Medline: 30484425]
15. Chan TM, Dzara K, Dimeo SP, Bhalerao A, Maggio LA. Social media in knowledge translation and education for physicians and trainees: a scoping review. Perspect Med Educ 2020 Feb;9(1):20-30 [FREE Full text] [doi: 10.1007/s40037-019-00542-7] [Medline: 31834598]
16. O'Connor C, DeCaire K, Friedrich J. Improving patient care through the use of evidence based order sets. AMIA Annu Symp Proc 2005 Jan 01:1063 [FREE Full text] [Medline: 16779350]
17. Bobb AM, Payne TH, Gross PA. Viewpoint: controversies surrounding use of order sets for clinical decision support in computerized provider order entry. J Am Med Inform Assoc 2007;14(1):41-47 [FREE Full text] [doi: 10.1197/jamia.M2184] [Medline: 17068352]
18. Wright A, Sittig DF, Carpenter JD, Krall MA, Pang JE, Middleton B. Order sets in computerized physician order entry systems: an analysis of seven sites. AMIA Annu Symp Proc 2010 Nov 13;2010:892-896 [FREE Full text] [Medline: 21347107]
19. McGreevey JD. Order sets in electronic health records: principles of good practice. Chest 2013 Jan;143(1):228-235. [doi: 10.1378/chest.12-0949] [Medline: 23276846]
20. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
21. Forseen SE, Corey AS. Clinical decision support and acute low back pain: evidence-based order sets. J Am Coll Radiol 2012 Oct;9(10):704-712. [doi: 10.1016/j.jacr.2012.02.014] [Medline: 23025864]

22. Charlotte W, Loshak H. Standardized hospital order sets in acute care: a review of clinical evidence, cost-effectiveness, and guidelines. In: CADTH Rapid Response Report: Summary with Critical Appraisal. Ottawa, Ontario, Canada: Canadian Agency for Drugs and Technologies in Health; Jul 25, 2019.
23. Ballard DJ, Ogola G, Fleming NS, Heck D, Gunderson J, Mehta R, et al. The impact of standardized order sets on quality and financial outcomes. In: Advances in Patient Safety: New Directions and Alternative Approaches. Rockville (MD): Agency for Healthcare Research and Quality (US); Aug 2008.
24. Hospital beds staffed and in operation, 2018-2019. Canadian Institute for Health Information. 2019. URL: https://www.cihi.ca/sites/default/files/document/beds-staffed-and-in-operation-2018-2019-en-web.xlsx [accessed 2020-06-25]
25. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
26. Straus SE, Tetroe J, Graham I. Defining knowledge translation. Can Med Asso J 2009 Aug 4;181(3-4):165-168 [FREE Full text] [doi: 10.1503/cmaj.081229] [Medline: 19620273]
27. Marchand-Senécal X, Kozak R, Mubareka S, Salt N, Gubbay JB, Eshaghi A, et al. Diagnosis and management of first case of COVID-19 in Canada: lessons applied from SARS-CoV-1. Clin Infect Dis 2020 Nov 19;71(16):2207-2210 [FREE Full text] [doi: 10.1093/cid/ciaa227] [Medline: 32147731]
28. Kupferschmidt K. 'A completely new culture of doing research.' Coronavirus outbreak changes how scientists communicate. Science. 2020 Feb 26. URL: https://www.science.org/news/2020/02/completely-new-culture-doing-research-coronavirus-outbreak-changes-how-scientists [accessed 2021-09-24]
29. Ghosh P, Schwartz G, Narouze S. Twitter as a powerful tool for communication between pain physicians during COVID-19 pandemic. Reg Anesth Pain Med 2021 Feb;46(2):187-188. [doi: 10.1136/rapm-2020-101530] [Medline: 32321859]
30. Chan AK, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Mar 30:1579-1582. [doi: 10.1111/anae.15057] [Medline: 32227594]
31. Helliwell JA, Bolton WS, Burke JR, Tiernan JP, Jayne DG, Chapman SJ. Global academic response to COVID-19: cross-sectional study. Learn Publ 2020 Jul 01:1317 [FREE Full text] [doi: 10.1002/leap.1317] [Medline: 32836910]
32. Glasser J. It's time for a new kind of electronic health record. Harvard Business Review. 2020 Jun 12. URL: https://hbr.org/2020/06/its-time-for-a-new-kind-of-electronic-health-record [accessed 2020-06-25]
33. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: more retweets for science-based information on coronavirus than for false information. Int Sociol 2020 Apr 15;35(4):377-392. [doi: 10.1177/0268580920914755]
34. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. Can J Emerg Med 2020 Jul;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]

Abbreviations

CME: continuing medical education
EHR: electronic health record
TCN: the collaborative network

Edited by R Kukafka; submitted 06.12.20; peer-reviewed by M Afzal, DP Ryan; comments to author 22.01.21; revised version received 07.06.21; accepted 05.07.21; published 29.11.21.

Please cite as:
Javidan AP, Brand A, Cameron A, D'Ovidio T, Persaud M, Lewis K, O'Connor C
Examination of a Canada-Wide Collaboration Platform for Order Sets: Retrospective Analysis
J Med Internet Res 2021;23(11):e26123
URL: https://www.jmir.org/2021/11/e26123
doi:10.2196/26123
PMID:34847055

Original Paper

Adherence With Online Therapy vs Face-to-Face Therapy and With Online Therapy vs Care as Usual: Secondary Analysis of Two Randomized Controlled Trials

Sonia Lippke[1], PhD; Lingling Gao[1], PhD; Franziska Maria Keller[1], MSc; Petra Becker[2], PhD; Alina Dahmen[2], MD, MA

[1]Department of Psychology and Methods, Jacobs University Bremen, Bremen, Germany

[2]Dr. Becker Klinikgruppe, Cologne, Germany

Corresponding Author:
Sonia Lippke, PhD
Department of Psychology and Methods, Jacobs University Bremen
Campus Ring 1
Bremen, 28759
Germany
Phone: 49 421200 ext 4730
Fax: 49 421200494730
Email: s.lippke@jacobs-university.de

Abstract

Background: Adherence to internet-delivered interventions targeting mental health such as online psychotherapeutic aftercare is important for the intervention's impact. High dropout rates limit the impact and generalizability of findings. Baseline differences may be putting patients at risk for dropping out, making comparisons between online with face-to-face (F2F) therapy and care as usual (CAU) necessary to examine.

Objective: This study investigated adherence to online, F2F, and CAU interventions as well as study dropout among these groups and the subjective evaluation of the therapeutic relationship. Sociodemographic, social-cognitive, and health-related variables were considered.

Methods: In a randomized controlled trial, 6023 patients were recruited, and 300 completed the baseline measures (T1), 144 completed T2 (retention 44%-52%), and 95 completed T3 (retention 24%-36%). Sociodemographic variables (eg, age, gender, marital status, educational level), social-cognitive determinants (eg, self-efficacy, social support), health-related variables (eg, depressiveness), and expectation towards the treatment for patients assigned to online or F2F were measured at T1.

Results: There were no significant differences between the groups regarding dropout rates (χ^2_1=0.02-1.06, P≥.30). Regarding adherence to the treatment condition, the online group outperformed the F2F and CAU conditions (P≤.01), indicating that patients randomized into the F2F and CAU control groups were much more likely to show nonadherent behavior in comparison with the online therapy groups. Within study groups, gender differences were significant only in the CAU group at T2, with women being more likely to drop out. At T3, age and marital status were also only significant in the CAU group. Patients in the online therapy group were significantly more satisfied with their treatment than patients in the F2F group (P=.02; Eta²=.09). Relationship satisfaction and success satisfaction were equally high (P>.30; Eta²=.02). Combining all study groups, patients who reported lower depressiveness scores at T1 (T2: odds ratio [OR] 0.55, 95% CI 0.35-0.87; T3: OR 0.56, 95% CI 0.37-0.92) were more likely to be retained, and patients who had higher self-efficacy (T2: OR 0.57, 95% CI 0.37-0.89; T3: OR 0.52, 95% CI 0.32-0.85) were more likely to drop out at T2 and T3. Additionally, at T3, the lower social support that patients reported was related to a higher likelihood of remaining in the study (OR 0.68, 95% CI 0.48-0.96). Comparing the 3 intervention groups, positive expectation was significantly related with questionnaire completion at T2 and T3 after controlling for other variables (T2: OR 1.64, 95% CI 1.08-2.50; T3: OR 1.59, 95% CI 1.01-2.51).

Conclusions: While online interventions have many advantages over F2F variants such as saving time and effort to commute to F2F therapy, they also create difficulties for therapists and hinder their ability to adequately react to patients' challenges. Accordingly, patient characteristics that might put them at risk for dropping out or not adhering to the treatment plan should be considered in future research and practice. Online aftercare, as described in this research, should be provided more often to medical rehabilitation patients.

Trial Registration: ClinicalTrials.gov NCT04989842; https://clinicaltrials.gov/ct2/show/NCT04989842

(J Med Internet Res 2021;23(11):e31274) doi:10.2196/31274

KEYWORDS

psychotherapeutic aftercare; medical rehabilitation; online therapy; face-to-face therapy; care as usual; retention; dropout

Introduction

Background

Internet-delivered, online interventions provide many advantages for the prevention and treatment of psychological problems and mental health disorders such as depression, anxiety, and functional limitation [1-3]. Psychotherapeutic online interventions have been shown to be effective [4,5]. However, 1 of 2 patients drop out from many online intervention studies [6]. Dropout limits the impact of these interventions and the generalizability of the findings [7]. Besides, few studies have compared a synchronous online therapy group guided by a therapist with a control group in a face-to-face (F2F) format using the same therapeutic concept ("Curriculum Hannover") and with a treatment/care as usual (CAU) group (eg, [8]).

So far, 1 study of the few pre-existing studies found an advantage of online therapy over CAU and about the same effects as the F2F format [9]. This research aimed to use the data from this previous research for a secondary data analysis to further investigate, within the German Pension Insurance's framework concept for rehabilitation therapy [10], adherence to the assigned treatment arm and patient dropout and for subjective evaluation of relationship, success, and satisfaction.

Adherence is a key concept [11,12] and is conceptualized as adhering to the assigned treatment (within this study) such as CAU. On the contrary, nonadherent behavior means that patients find themselves a (different) therapy to which they were not assigned: Nonadherence in the CAU arm means that patients registered in an F2F or online treatment format. Nonadherence in the F2F arm would mean that they chose another treatment outside the Curriculum Hannover treatment scope. Nonadherence in the online groups would mean that patients changed to a treatment other than online Curriculum Hannover. Consequences of the lack of instant availability of psychotherapists [13] and psychotherapy were found in terms of patients being nonadherent and joining other kinds of treatments such as inpatient psychotherapeutic treatment, drug therapy, or self-help groups. In a systematic review, lack of time was clearly related with lower adherence [14].

Dropout from the study in all study arms was conceptualized as not completing the questionnaire anymore because the patient intentionally left the study or was not reachable anymore for further measurements [6,7].

It is important to understand why patients display intervention adherence and study retention (as opposed to dropping out [3,6-8]): Different factors can predict whether patients remain in the assigned therapy and the study in general. If we know about such predictors, we can address them so that the program is nurturing the patients' needs better and to prevent dropout and nonadherence, with the resulting loss of intervention efficacy and effectiveness [5]. This can also improve the impact of the treatments [4]. Accordingly, in the following, the evidence regarding potential correlates and predictors are summarized to explain our study's design described in the methods section.

Prior Work on Study Dropout and Adherence

Expansive knowledge already exists on factors affecting study participants, the likelihood of questionnaire completion (eg, [6]), and adherence to the assigned treatment. For instance, in their systematic review of 33 randomized controlled trials (RCTs), Brown et al [11] identified the following reasons for low adherence levels: time issues, little or no interest of the participants, the perception of the participants that treatment is not needed at all or anymore, or the intervention is not effective. They also identified technical problems and other priorities in daily life including holidays and work. Moreover, dissatisfaction with the assigned group was shown to be important. Furthermore, health issues and a fading motivation to participate in the program were also found [11]. The authors also revealed no statistical relationship between the intended duration of the program and adherence with the intervention.

However, a direct comparison of completers in the different intervention arms such as online vs F2F is rare. Further, evidence regarding what drives patients to adhere to the assigned therapy is scarce within internet-delivered psychotherapeutic interventions. A previous study with students [12] investigated counseling delivered online or F2F versus a placebo treatment. Lack of motivation, dissatisfaction with the counseling process, and perceptions that the counselor would not understand students via this medium were all more prominent in online counseling than in F2F counseling [8,12]. However, physical appearance was indicated as a barrier in the F2F group but not in the online group [12]. Moreover, F2F counseling was perceived as excessively straining in light of other duties [12].

Advantages of Online Interventions

Past studies [4,5] have consistently found that online treatments can save the therapists time and support relapse prevention after F2F therapy. Additional strengths of online interventions over F2F interventions are that they are deliverable from remote locations, need less time commitment, and provide more flexibility for therapists and patients. Another advantage may be that the risk of stigma due to a mental disorder and seeking treatment is reduced [4]. This can overcome the problems with F2F therapies, which furthermore are often not readily available in all regions and where they are needed, resulting in patients promptly starting with their online intervention instead of waiting a long time (which is typical for F2F therapies due to limited availability of therapists [13]). Moreover, patients in psychotherapeutic interventions may miss their F2F sessions or drop out of therapy because they feel as if the location of the therapy is too far away [15]. Thus, online mental health

interventions can bridge the gap between patients and therapists when the patient cannot travel to the intervention site (eg, [7]) or both are limited in their mobility.

Disadvantages of Online Interventions

Online therapies may harbor weaknesses like the requirement of knowledge and skills such as computer and internet health literacy and general literacy [4]. Not every patient may benefit from online psychotherapy or blended therapy forms (ie, a combination of F2F psychotherapy with online interventions modes) due to limited introspection capabilities or the nature of their disorder (ie, severe disorders, chronic syndromes, or personality disorders [13]). Hence, a personalized tool may be needed to consider individual patient characteristics [4]. Traditional therapy settings can help patients with self-reflection, especially if they are not well experienced with expressing their cognitions and emotions [4]. Additionally, online therapies may prevent counselors from reacting to emergency situations like acute psychic decompensation or acute psychosis as adequately as they could in an analogue situation [4]. Negative experiences with digital psychotherapeutic interventions could have the consequence of patients feeling less motivated, feeling unsure, or even avoiding trying F2F therapy. Conversely, relative to F2F interventions, online interventions might have limitations such as higher dropout (eg, 43% computer-guided or online vs 24% clinician-guided or F2F [13]). However, the opposite pattern was found in studies (ie, lower dropout in online interventions [4]). Thus, this needs more systematic investigation.

Sociodemographic Correlates

In some studies, age was found to be related to the willingness to participate and remain in online research [14], with younger individuals having a higher likelihood of participation in general, but also dropping out easier, than older individuals [16,17].

Women are typically more likely to participate in surveys (eg, [12]), but this is often reversed when internet studies are reviewed (eg, [16]). It is explained by the finding that men have, on average, more favorable attitudes, beliefs, and self-efficacy expectations toward technology use [18]. With that, men are more likely to remain in some online intervention studies, as indicated by a higher completion rate of modules compared with women [14], and more likely to drop out in other studies (eg, [19]).

Marital status, social integration, and social support are helpful for retaining patients in online interventions [8,14,17] and to mediate the intervention effect on symptomatology [7,16]. Study participants with low social support were more likely to seek such social support or consumer feedback in treatments [14,20]. Studies also demonstrated that individuals who heard about the online treatment from a family member and those with social stress were more likely to be strained and to drop out from the treatment. Dropout for these individuals was largely due to family responsibilities such as caring for a child or another family member [8,17,20].

The typical finding regarding educational level is that more literate participants harbor a higher willingness to participate in and are more successful at remaining in the study [20,21]. On the contrary, individuals with poor education, low income, and higher risk for unemployment were more likely to not make use of health care or to not benefit from it compared with their more privileged counterparts [16,17]. However, an aggregation of only online psychological interventions revealed no conclusive evidence regarding employment [14].

Health-Related, Psychosocial, and Social-Cognitive Variables

Regarding health-related and psychosocial variables, along with workability, the typical pattern is that the more somatoform complaints, depressive symptoms, and phobic anxiety, the lower their mental well-being; in addition, the more interactional difficulties patients have, the more likely they are to drop out [6,17,20].

Regarding social-cognitive variables, self-efficacy was found not to be directly related to adherence but to planning, which predicted adherence [3,14]. Finally, higher treatment expectancy regarding the treatment efficacy was related to greater study retention and adherence because improvement expectation also helped participants overcome difficulties and remain motivated [14,20]. These findings can be explained using the theoretical backdrop of the health action process approach (HAPA, [22]), which will also be used for this study. In the study by Zarski et al [3], 14% of the variance in treatment adherence could be explained by variables in the HAPA model; however, it remains unknown whether other characteristics beyond the HAPA variables can explain more variance in study dropout.

Therapeutic Relationship

For any form of psychosomatic rehabilitation aftercare led by a therapist, establishing a therapeutic relationship is of tremendous importance [7]. To successfully provide online therapies and build a therapeutic relationship, therapists require training, induction, clear guidelines, in-depth information and training, and continuing education and training [23].

This leads to the following question: If the therapists are well-trained in computer literacy and internet skills, will satisfaction with the relationship and patient progress in therapy be equal in the online therapy and the F2F treatment?

Goal of This Study

Due to the various sociodemographic variables, social-cognitive determinants, and health-related variables influencing adherence to online and F2F-therapies, we need to consider whether baseline differences may be putting patients at risk for dropping out.

This study made use of a setting with medical rehabilitation and aftercare [10]. Medical rehabilitation for psychosomatic patients in Germany is provided in clinics, where disadvantages of online interventions can also be addressed to enable patients and therapists as well to use online psychosomatic aftercare. For instance, technical requirements can be cleared, and affinity for internet use (ie, digital health literacy) can be trained to ensure patient-treatment fit [8].

However, so far, no study can be found addressing the comparison of the same aftercare delivered online (content and

procedure of the therapy; Curriculum Hannover [24,25]) with F2F and CAU in terms of study dropout and intervention withdrawal. Thus, this was the main aim of this study, since the usefulness of Curriculum Hannover was tested before and clearly revealed its superiority to CAU, as well as parity between the internet and F2F delivery modes [9].

Research Questions

We aimed to answer the following research questions: (1) What difference exists between online therapy groups (ONL1 [equivalence study], ONL2 [superiority study]) and control groups (F2F, CAU) regarding patients who complete the questionnaires (completers) and patients who adhere to the assigned therapy? (2) What reasons do patients have for not adhering with the assigned therapy? (3) What factors are related with completing the different study arms at T2 and T3? (4) Does the improvement expectation differ between the online therapy groups (ONL1, ONL2) and the control groups (F2F, CAU), and does it relate to questionnaire completion at T2 (9 [superiority study] or 12 [equivalence study] months after the end of the rehabilitation) and T3 (15 [superiority study] or 18 [equivalence study] months after the end of rehabilitation) after controlling for other variables? (5) Does the subjective evaluation of the online psychotherapy (ONL1, ONL2) and F2F psychotherapy differ with regard to relationship satisfaction, success satisfaction, and satisfaction with therapy?

Methods

Ethical Considerations

All participants were informed about the purpose of the study (including information on the length of the questionnaires and data storage procedures) via a participant information form and an informed consent form (all forms can be seen in the appendix to the CONSORT eHealth statement [9]). All procedures conducted in this study were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The questionnaire prior to the intervention was mandatory for every study participant to avoid missing units. However, whether study participants actually answered the individual questions was voluntary. In the case of questions from the patient, a project manager was at hand to reduce the risk of dropout from the study. The study protocol was published together with the primary results [9].

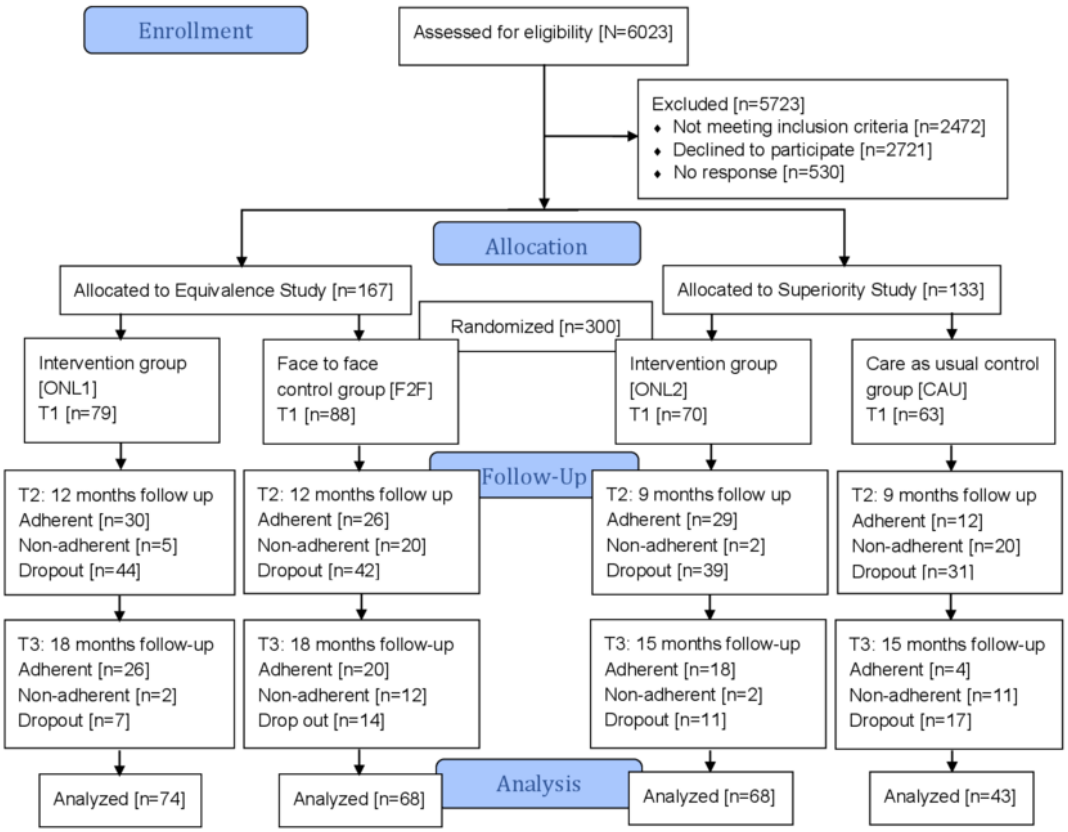
The study protocol was approved by the Ethics Committee of the North Rhine Medical Association (Ärztekammer Nordrhein; No. 2015351; Dec 4, 2015). As this study was run in a rehabilitation treatment and aftercare setting, approval was given by the pension funds (Bund, Braunschweig-Hannover and Rheinland) and corresponding patient councils [9]. The clinical trial was retrospectively registered with ClinicalTrials.gov (NCT04989842).

Recruitment

We recruited 6023 patients at their psychosomatic rehabilitation clinic. After excluding noneligible patients, 300 completed the baseline measures (T1). All rehabilitants who had participated in a psychotherapeutic rehabilitation treatment with Dr. Becker Klinik Möhnesee (March 2017 to May 2018), Dr. Becker Klinik Juliana (March 2017 to April 2018), or Dr. Becker Burgklinik (complete period between March 2017 and September 2020) were eligible for the aftercare therapy offered following the medical rehabilitation program [9].

Rehabilitants were questioned during their stay by a member of the social services staff and asked whether they wanted to take advantage of an offer for aftercare therapy. If they said yes, they were informed about the option of participating in the study and about the study conditions. If they agreed to participate, the rehabilitants who had a potential F2F therapy offering within a 45-minute radius of their place of residence were randomly assigned to either F2F therapy or online therapy within the *equivalence study* (Figure 1) [9]. Those without therapy offerings in their vicinity were randomly assigned to online therapy or no therapy within the so-called *superiority study*.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart for the Curriculum Hannover. CAU: care as usual, F2F: face-to-face, ONL: online.



Overall, 300 rehabilitants were assigned to one of the study arms: For the equivalence study, 167 rehabilitants were randomized to the online therapy (ONL1) or the F2F therapy (Figure 1). For the superiority study, 133 rehabilitants were randomly assigned to online therapy (ONL2) or the control group (CAU) [9].

Study Participants

After providing informed consent, 300 patients were included in the study (see Figure 1). All patients had participated in rehabilitation treatment [10] lasting between 1 week and 8 weeks (median 5 weeks; mean 5.11, SD 1.30 weeks) in 1 of 4 clinics [9].

Individuals had been admitted to rehabilitative treatment due to psychosomatic diagnoses according to the International Classification of Disease, 10th revision (ICD-10) manual (F20.0-F61 and G43.1), with the most frequent diagnoses being F32.1 Depressive episode (55/300), F33.1 Recurrent depressive disorder (87/300), and Adjustment disorders (33/300). At the conclusion of their rehabilitation, all individuals were encouraged to maintain their rehabilitation treatment success [9,10,24,25]. Individuals received the following treatment recommendations: medication check-up (270/300); psychotherapy and psychological counseling (246/300); therapeutic support including physio-, ergo-, or dietary therapy (171/300); aftercare (162/300); addiction counseling (32/300); stationary treatment (4/300); hours-wise reintegration into the job (31/300); diagnostics (30/300); return to work support (21/300); self-help group (5/300); and other recommendations (180/300).

The sample had a mean age of 50.27 (SD 9.71; range 25-67; median 52; mode 58) years. More women participated in the study (199/300, 66.3%), 51.0% (153/300) were partnered/married, and 75.7% (227/300) were employed.

Implementation of Therapy

The therapy in the usual F2F contact was based on the concept from Curriculum Hannover [24,25], an aftercare treatment following the medical rehabilitation. Apart from the admission and final interviews, which were each a 50-minute individual interview, therapy took place over 25 weekly 90-minute group sessions with 8-10 participants.

The online aftercare was carried out according to the same concept but had technical peculiarities due to the digital format [9,10,24,25]: The participants were instructed in advance on how to use the video platform by means of a learning video; this included, among other things, the rules for communication in the virtual group room as well as instructions for regularly checking the internet connection. The psychotherapists prepared for the special features of the new format through training courses geared to equip them to build a therapeutic relationship in a targeted manner. Also, they were instructed to teach their patients how to use the video platform. As a part of the group therapy, psychoeducation was conveyed using PowerPoint presentations or a whiteboard. Handouts or homework could also be distributed [9].

Under study conditions, CAU therapy meant that no standardized therapy measures were initiated. For these patients, only whether they independently took up treatments was assessed.

Statistical Analysis

Comparisons

Differences between the intervention and control groups in terms of dropout and adherence (research question 1) were tested with frequency analyses (chi-squared). Patients' reasons as to why they did not adhere (research question 2 and the sample description) were examined using descriptive analyses without calculating any statistics.

To test what factors were related to completing the different study arms (research question 3), *t* tests, chi-squared tests, and logistic regression analyses (determining odds ratios [ORs]) were conducted.

A multiple analysis of variance (MANOVA) with post-hoc Bonferroni tests was performed testing whether the subjective evaluation of the online and F2F therapy would differ with regard to relationship satisfaction, success satisfaction, and satisfaction with therapy (research question 4). This was also followed up by logistic regression analyses.

To test whether the subjective evaluation of the online and F2F therapy differed with regard to relationship satisfaction, success satisfaction, and satisfaction with therapy, a MANOVA was performed (research question 5) taking expectations into account. All analyses were run using SPSS 26.

Power

To determine the minimum required sample size through a priori analysis for obtaining a significant medium effect size, we used PADD 11 software and G*Power 3 software (Psychonomic Society Inc, Düsseldorf, Germany) [26,27]. With regard to the superiority study and based on the assumption of an effect of the intervention with a Cohen d>0.60 (>8 points difference in the primary endpoint), a total of 90 subjects (45 per study arm) was needed to conduct the study at an alpha <.05 and power >0.80. Concerning the equivalence study and based on the assumption of an equivalence margin of Cohen d<0.29 (<4 points difference in the primary endpoint), a total of 410 subjects (205 per study arm) was planned for the study to test at an alpha <.05 and power >.80 [24,25].

Data Exclusion

Inclusion criteria were an indication for psychosomatic therapy (indication is determined by the procedure described in the German Pension Insurance's framework concept for rehabilitation therapy [10]) and access to a standard PC, tablet, or smartphone with internet access (DSL or LTE) [9].

The exclusion criteria were also based on the framework concept for rehabilitation therapy of the German Pension Insurance and included employability <3 hours/day on the general labor market, drawing or applying for an old-age pension of at least two-thirds of the full-time pension, and drawing a benefit that is paid regularly until the start of a pension due to old age. Furthermore, patients with acute psychosomatic disorders were excluded [9,10].

The survey was conducted during rehabilitation by means of questionnaire measures at baseline (T1), 9 (superiority study) or 12 (equivalence study) months after the end of rehabilitation (after completion of the therapy intervention; T2), and 15 (superiority study) or 18 (equivalence study) months after the end of rehabilitation (T3). The reason for the postponed survey in the equivalence study was the wait time for an aftercare place in the F2F therapy.

The study participants were asked to fill out the T2 and T3 questionnaires by email and were reminded 2 weeks and 4 weeks later, respectively. The questionnaires were filled out digitally via the platform soscisurvey.de. At T2 and T3, the patients were asked whether they really followed the study protocol in terms of adhering to the study arm and its treatment as they were assigned [10].

Survey Instruments

The following items were analyzed in this longitudinal study [9]: sociodemographic information on gender, age, marital status, educational level, employment status, and income level.

To measure different aspects of mental health, the Hamburg Modules for the Assessment of Psychosocial Health in Clinical Practice (HEALTH-49) [28] were used for the purpose of this study. Individuals were asked to answer different modules, such as their feelings during the past 2 weeks or symptoms during the past 2 weeks, assessing various aspects of mental health. Items belonging to modules A and C were measured on a 5-point Likert scale from 1 ("not at all") to 5 ("very much"). Modules B, E, and F were also measured on a 5-point Likert scale from 1 ("never") to 5 ("always"). Module D was also measured on a 5-point Likert scale from 1 ("not true") to 5 ("very true").

Subjective employment prognosis [29] was measured via 3 items. The first item assessed whether individuals believed they would work until retirement. This item was measured on a 5-point Likert scale from 1 ("sure") to 5 ("not at all"). The second item assessed whether individuals perceived their overall earning capacity was at risk due to their health status. This item was measured on a 2-point Likert scale from 1 ("no") to 2 ("yes"). The third and last item evaluated whether individuals were thinking about applying for pension due to health limitations, and this was measured on a 3-point Likert scale: 1 ("no"), 2 ("yes"), and 3 ("I have already submitted a pension application").

In addition, 3 items of the Work Ability Index (WAI) [30] were implemented to provide an assessment of the ability to work. The first item, "If you rate your best ever work ability as 10 points: How many points would you then give for your current work ability (0 means you are currently unable to work)?" was assessed on an 11-point Likert scale from 0 ("completely unable to work") to 10 ("currently the best working capacity"). In addition, the 2 items "How would you rate your current work ability in terms of physical job demands?" and "How would you rate your current work ability in terms of mental work demands?" were both measured on a 5-point Likert scale from 1 ("very poorly") to 5 ("very good").

Additionally, the subscale hope of improvement of the standardized Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV) was used to assess their expectation of improvement in symptoms after therapy [31]. Participants were asked 4 items, which were measured on a 5-point Likert scale from 1 (“completely disagree”) to 5 (“completely agree”).

Further, the therapeutic relationship was assessed by means of the Helping Alliance Questionnaire (HAQ) using 2 subscales [32]: Subscale 1 focuses on relationship satisfaction (8 items), and subscale 2 focuses on success satisfaction (3 items). All items were measured on a 6-point Likert scale from 1 (“very inaccurate”) to 6 (“very accurate”).

Satisfaction with rehabilitation was evaluated using a standardized 8-item questionnaire [33]. Items were assessed on a 5-point Likert scale with different anchors used for the 8 items. Additionally, a self-developed questionnaire was used to assess participation in outpatient measures and therapy in the CAU group.

Results

Differences Between Groups

The 300 participants were divided into 2 RCTs: equivalence study (n=167) and superiority study (n=133; Table 1). All 300 participants completed the questionnaire at baseline (T1). The number of participants who completed the questionnaires at T2 and T3 is shown in Table 1 indicating retention rates.

Table 1. Study retention rates for the follow-up measurements in the equivalence study (online therapy vs face-to-face; n=167) and the superiority study (online therapy vs care as usual; n=133).

Time point	Equivalence study				Superiority study			
	ONL1[a] (n=79), n (%)	F2F[b] (n=88), n (%)	χ^2_1	*P* value	ONL2[c] (n=70), n (%)	CAU[d] (n=63), n (%)	χ^2_1	*P* value
T2[e]	35 (44)	46 (52)	1.06	.30	31 (44)	32 (51)	0.56	.453
T3[f]	28 (35)	32 (36)	0.02	.90	20 (29)	15 (24)	0.39	.533

[a]ONL1: online therapy in the equivalence study.

[b]F2F: face-to-face therapy.

[c]ONL2: online therapy in the superiority study.

[d]CAU: care as usual.

[e]T2: 12-month follow-up measurement for the equivalence study; 9-month follow-up measurement for the superiority study.

[f]T3: 18-month follow-up measurement for the equivalence study; 15-month follow-up measurement for the superiority study.

At T2, 2 participants partially completed the questionnaire and did not respond to questions regarding the outcome of treatment. These 2 patients were still labeled as completers as they responded to the questionnaire in general (and were included in all analyses except for the therapeutic relationship and satisfaction). All participants who completed the questionnaire at T3 also completed the questionnaire at T2. No differences between the ONL1 and F2F groups were found either at T2 or T3. Similarly, no differences between the ONL2 and CAU groups were found at either T2 or T3.

These results partially answer research question 1: There was no significant difference between the online therapy groups (ONL1, ONL2) and the control groups (F2F, CAU) regarding the percentage of patients who completed the questionnaires (identified as completers). To also test whether patients in the online therapy groups differed regarding their adherence to the assigned therapy, we further investigated the nonadherence rates across all patients at all time points. Differences per measurement point were tested within the 167 patients assigned to the equivalence study and the 133 patients in the superiority study (Table 2).

Table 2. Study retention (number and percentage of patients who completed the questionnaires; completers) and dropout rates (no completion of the questionnaires), as well as adherence to the assigned therapy for both the follow-up measurement points T2 and T3, by equivalence study (n=167) and superiority study (n=133).

Time points	Equivalence study				Superiority study			
	ONL1[a] (n=79), n (%)	F2F[b] (n=88), n (%)	χ^2_2	*P* value	ONL2[c] (n=70), n (%)	CAU[d] (n=63), n (%)	χ^2_2	*P* value
T2[e]								
Adherent completers	30 (38)	26 (30)	8.87	.012	29 (41)	12 (19)	22.38	<.001
Nonadherent completers	5 (6)	20 (23)			2 (3)	20 (32)		
Dropouts	44 (56)	42 (48)			39 (56)	31 (49)		
T3[f]								
Adherent completers	26 (33)	20 (23)	_g	_g	18 (26)	4 (6)	_g	_g
Nonadherent completers	2 (3)	12 (14)			2 (3)	11 (18)		
Dropouts	51 (65)	56 (64)			50 (71)	48 (76)		

[a]ONL1: online therapy in the equivalence study.

[b]F2F: face-to-face therapy in the equivalence study.

[c]ONL2: online therapy in superiority study.

[d]CAU: care as usual in the superiority study.

[e]T2: 12-month follow-up measurement for the equivalence study; 9-month follow-up measurement for the superiority study.

[f]T3: 18-month follow-up measurement for the equivalence study; 15-month follow-up measurement for the superiority study.

[g]Test statistic could not be computed because the cell frequencies were too small.

Patients randomized into the F2F and CAU control groups were much more likely to show nonadherent behavior (not adhering to the therapy to which the study participants were assigned) in comparison with the online therapy groups. In both studies, these differences in dropout and nonadherence were statistically significant (Table 2). More patients dropped out from the study at T3, and the nonadherence rates were lower than at T2 (Table 2). Descriptively, the highest percentage of study participants retained in the study were in the F2F group (32/88, 36%), whereas patients randomized to online therapy were retained at a slightly smaller percentage (ONL1: 28/79, 35% and ONL2: 20/70, 29%; Table 2). The highest risk for dropout from the study was in the CAU group (which did not receive any therapy within the study). However, the differences could not be analyzed statistically due to small cell sizes at T3.

Summarizing findings regarding research question 1, differences between the online therapy groups (ONL1, ONL2) and the control groups (F2F, CAU) occurred regarding both the percentage of patients who completed the questionnaires (completers) and patients who adhered to the assigned condition. These differences were statistically significant at T2 and only descriptive at T3.

Exploring Reasons for Nonadherence

To study research question 2, at T2, nonadherent patients were asked, using a set of predefined answers, why they did not adhere to the therapy to which they were assigned, and each patient could select the answers that most applied. Among patients randomized to ONL1 and identified as nonadherent (5/79), 2 patients answered that they felt the rehabilitation was already sufficient to meet their therapy goals and therefore did not see any further need to participate in the aftercare therapy. One patient indicated that s/he would not perceive the therapy as useful. One patient indicated a lack of motivation to attend the therapy, whereas another patient cited time constraints.

Of the patients randomized into the F2F group who were nonadherent (20/88), about one-third (7/20, 35%) cited the unavailability of such treatment, could not find therapy at all, or would have had to wait too long. There were 5 patients who answered that the location of the therapy would be too far away. Another 2 patients replied that they would not perceive the therapy as useful, while 2 patients indicated that the therapy would be too much of a strain. One patient indicated that s/he could not motivate her/himself to attend the therapy.

The 2 nonadherent participants from the 70 participants in the ONL2 group both answered that they did not participate in the therapy due to time constraints and other reasons. Of the 63 patients assigned to CAU, 20 were nonadherent: Of those, 16 patients participated independently from this study in an outpatient treatment. Another 2 underwent inpatient treatment, 14 received drug therapy, and 2 attended a self-help group.

Thus, to address research question 2, patients had different reasons for not adhering to the assigned therapy related to the study arm. Summarizing the online treatments (ONL1 and ONL2), the 7 nonadherent patients indicated no motivation (n=1), lack of understanding of the benefits of aftercare (n=1), no time (n=2), rehabilitation goals already achieved (n=2).

Comparisons Between Retained Patients and Those Who Dropped Out at T2 and T3

To investigate research question 3 and the differences between patients who completed the study and those who dropped out (not retained in the study), we compared sociodemographic

variables within the study arms at T2 (Table 3) and T3 (Multimedia Appendix 1). At T2, only gender was related with study dropout in the CAU group: While the completer group consisted of 59% (19/32) women, the dropouts were female (26/31, 84%) with a higheer likelihood. Thus, within the CAU group, it seemed to be more difficult for women to be retained in the study than it was for men. However, such gender differences did not appear in any other group at a significant level (see Table 3).

Table 3. The differences between patients who completed the study and those who dropped out at T2 (12-month follow-up measurement for the equivalence study; n=167; 9-month follow-up measurement for the superiority study; n=133).

Variables	Equivalence study ONL1[a]			Equivalence study F2F[b]			Superiority study ONL2[c]			Superiority study CAU[d]		
	Completers[e] (n=35)	Dropouts[f] (n=44)	*P* value	Completers[e] (n=46)	Dropouts[f] (n=42)	*P* value	Completers[e], (n=31)	Dropouts[f] (n=39)	*P* value	Completers[e] (n=32)	Dropouts[f] (n=31)	*P* value
Age (years), mean (SD)	50.23 (9.36)	50.98 (9.09)	.72	51.48 (8.88)	49.50 (10.26)	.34	51.32 (10.15)	48.97 (10.36)	.35	51.75 (9.32)	47.58 (10.65)	.10
Gender (female), n (%)	27 (77)	26 (59)	.09	27 (59)	24 (57)	.88	19 (61)	31 (80)	.09	19 (59)	26 (84)	.031
Marital status (Married), n (%)	24 (67)	25 (57)	.29	23 (50)	23 (55)	.66	16 (52)	17 (44)	.50	18 (56)	11 (36)	.10
Educational level, n (%)								.				
Elementary school	18 (51)	23 (52)	.90	19 (41)	22 (52)	.58	16 (52)	26 (67)	.24	16 (50)	11 (36)	.38
High school	6 (17)	10 (23)		14 (30)	10 (24)		2 (7)	4 (10)		5 (16)	9 (29)	
College and above	9 (26)	9 (21)		13 (28)	10 (24)		13 (42)	9 (23)		11 (34)	10 (32)	
Other	2 (7)	2 (5)		0	0		0	0		0	1 (3)	
Employment status (employed), n (%)	27 (77)	34 (77)	.99	34 (74)	29 (69)	.61	27 (87)	30 (77)	.28	23 (72)	23 (74)	.84
Income status (€[g]), n (%)												
<1500	8 (23)	11 (25)	.52	13 (28)	15 (36)	.65	11 (36)	13 (33)	.38	9 (28)	9 (29)	.67
1500-3000	17 (49)	16 (36)		23 (50)	17 (41)		9 (29)	17 (44)		15 (47)	17 (55)	
>3000	10 (29)	17 (39)		10 (22)	10 (24)		11 (36)	9 (56)		8 (25)	5 (16)	

[a]ONL1: online therapy in the equivalence study.

[b]F2F: face-to-face therapy in the equivalence study.

[c]ONL2: online therapy in the superiority study.

[d]CAU: care as usual in the superiority study.

[e]Completers: study participants who completed the questionnaires.

[f]Dropouts: study participants who dropped out from the study and did not complete the questionnaires.

[g]A currency exchange rate of €1=US $1.16 is applicable.

To investigate research question 3 regarding T3, the same analyses were performed with the second follow-up measurement (Multimedia Appendix 1). Age and marital status emerged as significantly differentiating between completers and dropouts in the CAU group: In the completers group, patients were, on average, 7 years older than patients in the dropout group. Moreover, for patients randomized into the CAU group, married patients were more likely to be retained in the study, whereas single patients were much more likely to drop out. However, similar age and marital status differences did not appear on a significant level in any other group (see Multimedia Appendix 1). To summarize the results on research question 3, the only factors that were related to the completion of the different study arms at T2 and T3 were age, gender, and marital status. However, these were only bivariate analyses, and in the following sections, we investigate the interrelations in a multivariate approach.

Predicting Study Retention in All 4 Groups

To further examine research question 3 and to investigate whether the intervention group (Model 0), sociodemographic variables (Model 1), health-related and psychosocial variables, as well as work ability at baseline (Model 2) were interrelated with the completion of the study at T2 and T3 (Table 4), logistic regression analyses were performed. This was done using dummy coding for patients who remained in the study at T2 and T3 as 1 vs those who dropped out as 0.

After matching the results from the comparisons between the 4 groups, the study arm (ie, intervention group to which the patients were randomly assigned) was not significantly related with retention when comparing the 3 groups receiving a treatment (ONL1, F2F, ONL2) with the CAU control group (Intervention [ONL vs F2F] in Multimedia Appendix 2). After controlling for the sociodemographic variables (Model 2 T2, Model 2 T3, with no variable being significant), 2 characteristics emerged for study retention at T2 and T3: Patients who reported lower depressiveness scores at T1 (better mental health; T2: OR 0.51, 95% CI 0.30-0.85; T3: OR 0.45, 95% CI 0.25-0.79) were more likely to be retained in the study, and patients with higher self-efficacy (T2: OR 0.50, 95% CI 0.30-0.83; T3: 0.57, 95% CI 0.33-0.99) were more likely to drop out. Additionally, at T3, higher social support seemed to make it more likely that patients dropped out: the less social support patients reported, the more they remained in the study (but this was not significant at T2; T2: OR 0.79, 95% CI 0.54-1.15; T3: OR 0.67, 95% CI 0.44-0.99; Table 4).

These findings from the 3 intervention groups (Multimedia Appendix 2) matched the findings from all 4 groups (Table 4). Remarkably, with all 4 groups, being married emerged as significant in Model 2 when controlling for the health-related variables, psychosocial variables, and work ability at baseline, indicating a suppressor effect: Only when parts of the variance were explained by depressiveness, self-efficacy, and social support, those who were single were more likely to remain in the study. Additionally, on a marginal/descriptive level, findings from these analyses were validated, with retention higher in the F2F group (Model 0 T2: OR 1.06, 95% CI 0.56-2.03; Model 1 T2: OR 1.05, 95% CI 0.54-2.05; Model 2 T2: OR 1.14, 95% CI 0.57-2.28; Model 1 T3: OR 1.93, 95% CI 0.92-4.07; Model 2 T3: OR 2.10, 95% CI 0.96-4.59), for older participants (Model 1 T2: OR 1.02, 95% CI 0.99-1.04; Model 2 T2: OR 1.02, 95% CI 0.99-1.04; Model 1 T3: OR 1.03, 95% CI 0.99-1.06; Model 2 T3: OR 1.03, 95% CI 1.00-1.06), and for married participants (Model 1 T2: OR 1.55, 95% CI 0.94-2.57; Model 2 T2: OR 1.80, 95% CI 1.05-3.08; Model 1 T3: OR 1.47, 95% CI 0.85-2.52; Model 2 T3: OR 1.78, 95% CI 1.00-3.20; Table 4).

Summarizing findings regarding research question 3, the following factors were (partially) related with the completion of the different study arms at T2 and T3: age, marital status, depressiveness, self-efficacy, and social support.

Table 4. Logistic regression models predicting study retention with all patients: online therapy in the equivalence study (ONL1), face-to-face therapy in the equivalence study (F2F), online therapy in superiority study (ONL2), care as usual in the superiority study (CAU).

Variables	Model 0 T2[a]		Model 1 T2		Model 2 T2		Model 1 T3[b]		Model 2 T3	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Intervention group										
ONL1 vs CAU	0.77 (0.39-1.50)	.44	0.75 (0.38-1.50)	.42	0.74 (0.37-1.52)	.43	1.79 (0.83-3.86)	.14	1.82 (0.82-4.03)	.14
F2F vs CAU	1.06 (0.56-2.03)	.86	1.05 (0.54-2.05)	.88	1.14 (0.57-2.28)	.71	1.93 (0.92-4.07)	.08	2.10 (0.96-4.59)	.06
ONL2 vs CAU	0.77 (0.39-1.52)	.45	0.75 (0.37-1.53)	.43	0.78 (0.38-1.64)	.52	1.29 (0.58-2.88)	.54	1.32 (0.57-3.07)	.52
Age	N/A[d]	N/A	1.02 (0.99-1.04)	.21	1.02 (0.99-1.04)	.28	1.03 (0.99-1.06)	.06	1.03 (1.00-1.06)	.08
Gender										
Female	N/A	N/A	1	N/A	1	N/A	1	N/A	1	N/A
Male	N/A	N/A	1.15 (0.70-1.90)	.58	1.22 (0.72-2.07)	.45	1.00 (0.58-1.71)	.99	1.01 (0.57-1.78)	.98
Marital status										
Unmarried	N/A	N/A	1	N/A	1	N/A	1	N/A	1	N/A
Married	N/A	N/A	1.55 (0.94-2.57)	.09	1.80 (1.05-3.08)	.03	1.47 (0.85-2.52)	.17	1.78 (1.00-3.20)	.05
Educational level										
Elementary school	N/A	N/A	1	N/A	1	N/A	1	N/A	1	N/A
High school	N/A	N/A	0.89 (0.48-1.66)	.72	1.02 (0.53-1.98)	.95	0.85 (0.43-1.70)	.65	0.96 (0.46-2.00)	.91
College and above	N/A	N/A	1.42 (0.81-2.49)	.22	1.63 (0.90-2.96)	.11	1.28 (0.70-2.35)	.42	1.50 (0.79-2.85)	.22
Other	N/A	N/A	0.75 (0.12-4.82)	.76	0.68 (0.10-4.62)	.70	2.23 (0.18-8.19)	.83	1.07 (0.15-7.78)	.95
Employment status										
Unemployed	N/A	N/A	1	N/A	1	N/A	1	N/A	1	N/A
Employed	N/A	N/A	1.13 (0.62-2.05)	.70	1.13 (0.60-2.12)	.71	1.51 (0.77-2.95)	.23	1.68 (0.82-3.45)	.16
Income status (€[e])										
<1500	N/A	N/A	1	N/A	1	N/A	1	N/A	1	N/A
1500-3000	N/A	N/A	0.96 (0.53-1.76)	.91	0.91 (0.48-1.72)	.78	1.04 (0.54-1.99)	.92	1.06 (0.53-2.12)	.87
>3000	N/A	N/A	0.76 (0.37-1.56)	.45	0.78 (0.36-1.66)	.51	0.58 (0.26-1.28)	.18	0.63 (0.27-1.46)	.28
Somatoform complaints	N/A	N/A	N/A	N/A	0.97 (0.67-1.38)	.84	N/A	N/A	0.91 (0.61-1.35)	.64
Depressiveness	N/A	N/A	N/A	N/A	0.55 (0.35-0.87)	.01	N/A	N/A	0.56 (0.37-0.92)	.02
Phobic fear	N/A	N/A	N/A	N/A	1.15 (0.81-1.64)	.42	N/A	N/A	1.30 (0.88-1.93)	.18
Mental well-being	N/A	N/A	N/A	N/A	1.47 (0.90-2.40)	.13	N/A	N/A	1.63 (0.95-2.80)	.07
Interactional difficulties	N/A	N/A	N/A	N/A	1.21 (0.87-1.68)	.26	N/A	N/A	1.29 (0.90-1.86)	.16

Variables	Model 0 T2[a]		Model 1 T2		Model 2 T2		Model 1 T3[b]		Model 2 T3	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Self-efficacy	N/A	N/A	N/A	N/A	0.57 (0.37-0.89)	.01	N/A	N/A	0.52 (0.32-0.85)	.01
Activity and participation	N/A	N/A	N/A	N/A	1.25 (0.85-1.83)	.26	N/A	N/A	1.06 (0.70-1.60)	.80
Social stress	N/A	N/A	N/A	N/A	0.85 (0.60-1.21)	.37	N/A	N/A	0.71 (0.48-1.06)	.09
Social support	N/A	N/A	N/A	N/A	0.79 (0.58-1.08)	.14	N/A	N/A	0.68 (0.48-0.96)	.03
Work ability index	N/A	N/A	N/A	N/A	0.98 (0.76-1.27)	.89	N/A	N/A	0.96 (0.72-1.27)	.75
Nagelkerke	.007	.646	.043	.62	.119	.05	.067	.272	.197	.004

[a]T2: 12-month follow-up measurement for the equivalence study; 9-month follow-up measurement for the superiority study.

[b]T3: 18-month follow-up measurement for the equivalence study; 15-month follow-up measurement for the superiority study.

[c]OR: odds ratio.

[d]N/A: not applicable.

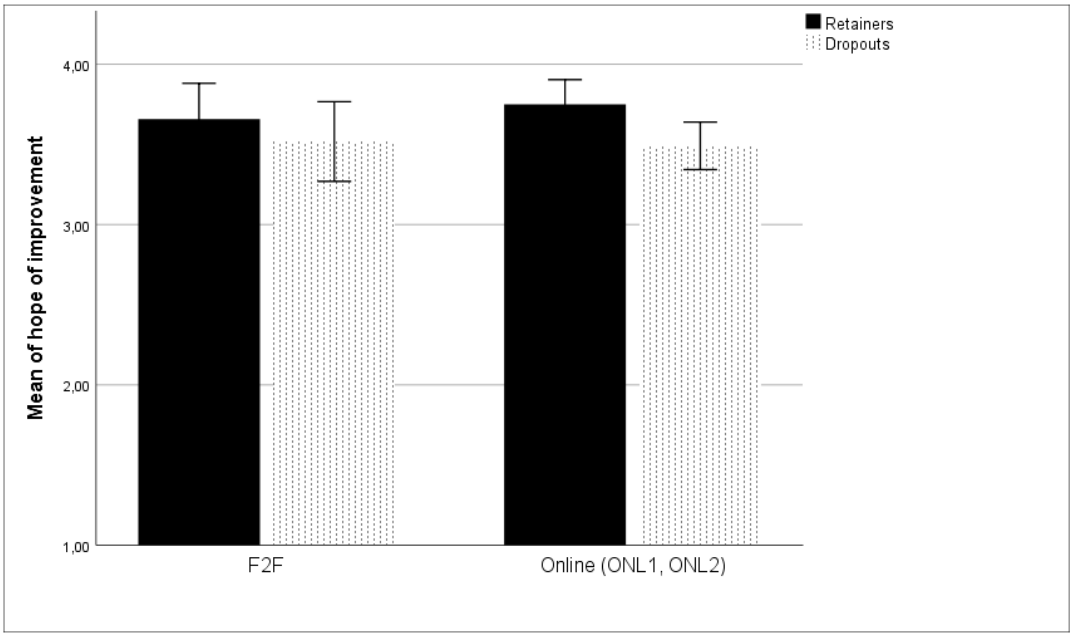
[e]A currency exchange rate of €1=US $1.16 is applicable.

Expectations for Improvement From Face-to-Face and Online Therapy

To test research question 4, improvement expectation from therapy was measured at baseline among the ONL1, F2F, and ONL2 groups to evaluate any differences. We conducted a MANOVA controlling for age, gender, and marital status. The results showed significant differences in improvement expectation based on the study arm ($F_{2,228}$=3.13, *P*=.05, Eta²=.027) and based on dropout rates ($F_{1,228}$=5.68, *P*=.02, Eta²=.024), after controlling for age (*P*=.96; Eta²<.001), gender (*P*=.89; Eta²<.001), and marital status (*P*=.91; Eta²<.001).

When conducting post-hoc comparisons of differences between patients who completed the questionnaires and those who did not at T2, the 2 online groups (ONL1+ONL2) were contrasted with the F2F group. We found a significant difference ($F_{1,147}$=5.74, *P*=.02): The mean score for improvement expectation from therapy among the online therapy groups was significantly higher among patients who completed the questionnaire than among those who did not complete the questionnaire at T2 (see Figure 2). The difference among the F2F therapy group did not show any significant effects ($F_{1,86}$=0.716, *P*=.40).

Figure 2. Mean scores for improvement expectation during therapy among patients who completed the questionnaires and those who dropped out at T2, in the face-to-face (F2F) and combined online groups (ONL1+ONL2).



When further exploring the post-hoc comparisons on improvement expectation between patients who completed the questionnaires and those who dropped out at T3 among online (ONL1+ONL2) vs F2F therapy, the combined online therapy group emerged as significant ($F_{1,147}$=5.74, *P*=.04). The mean score for improvement expectation was higher among patients who completed the questionnaire (mean 3.76, SD 0.61) than among patients who dropped out (mean 3.52, SD 0.68). The differences among patients receiving F2F therapy was not significant ($F_{1,83}$=0.486, *P*=.49).

Predicting Study Retention With Improvement Expectations in the Intervention Groups

Logistic regression analyses were performed as described with patients in the ONL1, F2F, and ONL2 groups and including improvement expectation (Multimedia Appendix 2). Improvement expectation was significant when included in the analyses: Those who had higher improvement expectations were also those who were more likely to complete study questionnaires (Model 2 T2: OR 1.64, 95% CI 1.08-2.50; Model 2 T3: OR 1.59, 95% CI 1.01-2.51; Multimedia Appendix 2).

To answer research question 4, we can summarize that improvement expectations did differ between the online therapy groups (ONL1, ONL2) and the control groups (F2F, CAU). Positive expectation was also significantly related with completion of the questionnaire at T2 and T3 after controlling for other variables: the more the patients in all groups expected, the more likely they were to be retained in the study.

Evaluation of the Different Therapies With Regard to Different Satisfaction Aspects

To test research question 5, whether subjective evaluations of the online and F2F therapies differed with regard to relationship satisfaction, success satisfaction, and satisfaction with therapy, we conducted a MANOVA. The therapeutic relationship included 2 aspects: relationship satisfaction (HAQ1) and success satisfaction (HAQ2). Both items, together with satisfaction with the treatment, were analyzed using a MANOVA to evaluate overall effects, and then we tested for group differences (study arm ONL1, F2F, ONL2, and patients with high vs low expectations) using Bonferroni post-hoc tests. Only the intervention groups explained significant proportions of the variance ($F_{\text{Roy's Largest Root};3,76}$=2.665, *P*=.05, Eta²=.095). Expectation and the interaction of intervention and expectation were not significant (*P*>.45). The means and SDs as well as summary statistics are reported in Table 5 along with the *F* tests of the individual test variables, indicating that the overall effect was based on satisfaction with the therapy. Within this variable, the post-hoc tests revealed that differences between groups only existed between online and F2F groups, but not between the 2 online groups.

Table 5. Statistical results of the therapeutic relationship and satisfaction with the therapy in patients with low expectations (below a median of 3.50, n=31) vs high expectations (above the median, n=52).

Variables	ONL1[a] (n=30), mean (SD)	F2F[b] (n=24), mean (SD)	ONL2[c] (n=29), mean (SD)	CAU[d]	F_{group}	*P* value	Eta²
Relationship satisfaction (HAQ1[e])							
Low expectations	2.63 (0.67)	2.84 (0.95)	2.52 (1.01)	N/A[f]	0.937	.40	.02
High expectations	2.90 (1.67)	2.98 (1.11)	2.38 (1.18)				
Success satisfaction (HAQ2)							
Low expectations	2.45 (0.61)	2.65 (0.83)	2.51 (1.18)	N/A	0.697	.50	.018
High expectations	2.46 (1.60)	2.87 (1.26)	2.21 (1.17)				
Satisfaction with therapy[g]							
Low expectations	3.16 (0.41)	2.80 (0.65)	3.50 (0.43)	N/A	3.907	.024	.092
High expectations	3.33 (0.94)	2.69 (1.06)	3.22 (0.91)				

[a]ONL1: online therapy in the equivalence study.

[b]F2F: face-to-face therapy in the equivalence study.

[c]ONL2: online therapy in the superiority study.

[d]CAU: care as usual in the superiority study.

[e]HAQ: Helping Alliance Questionnaire.

[f]N/A: not applicable because not measured in the CAU group.

[g]Effect for expectation was not significant. Post-hoc tests revealed a significant difference between satisfaction and therapy: ONL1 > F2F (*P*=.05); ONL2 > F2F (*P*=.03).

To summarize findings regarding research question 5, the subjective evaluation of the online and F2F therapies differed based on relationship satisfaction and success satisfaction and specifically with satisfaction with the therapy. Patients randomized into the online therapy were significantly more satisfied with their treatment than patients in the F2F group. Relationship satisfaction and success satisfaction were equally high in the online and the F2F treatments.

Discussion

In medical internet research, it has been shown that outpatient psychotherapeutic treatment after rehabilitation treatment is an important factor in ensuring the sustainability of treatment effects [1-9,24,25]. At the same time, research has demonstrated the benefits of online psychotherapy in comparison with F2F psychotherapy on site [34]. For the first time, this study showed the direct comparison of psychotherapeutic aftercare delivered online versus F2F with CAU at more than 1-year follow-up (ie, 15 months and 18 months after baseline) in the field of psychosomatic therapy regarding adherence and dropout rates. This is important due to the fact that individually tailored online and F2F therapies have greater interventional effects than standard therapy programs [3,5-9].

Principal Findings

Of 300 patients participating in the T1 measurement point, 167 were assigned to the equivalence study because they had F2F psychotherapeutic aftercare available, and 133 were assigned to the superiority study because of the unavailability of aftercare options. Within both groups, the patients were randomized to the online psychotherapeutic aftercare or the comparator group. Retention rates in the online groups were equal in all groups. However, retaining the patients for periods of 18 months (ONL1, F2F) and 15 months (ONL2, CAU) was rather difficult: While after 9 months or 12 months, 56% of the patients in the online therapy dropped out from the study, a further 9%-15% (ie, 65%-71% in total) had dropped out after 15 months or 18 months. Whether the difference in dropout between the F2F group (48%-64%) and the CAU group (49%-76%) was due to the longer time frame of the follow-up measurement point or to the different conditions remains unclear at this point.

In addition, nonadherence rates were tested, and the F2F and CAU groups were much more likely to show nonadherent behavior in comparison with the online therapy groups: While the online groups were only nonadherent at 3%-6%, the F2F group was much more likely to be nonadherent, at 14%-23%. Self-reported reasons were mainly unavailability of the treatment, waiting times that were too long, and a location that was too far away.

The highest nonadherence occurred in the CAU condition, at 18%-32%. Reported reasons for this nonadherence were that patients participated in an outpatient psychotherapeutic aftercare, inpatient psychotherapy, drug therapy, or a self-help group independently from this study. These alternatives to just keep waiting could, of course, be regarded as good for the patient and as functional behavior. In contrast, dropout appears to have been unfavorable, as the patients' condition may have worsened making them unable to participate in the study while experiencing no or inappropriate support. Thus, determining how to meet the needs of the patients is of the highest priority to prevent dropout and dysfunctional nonadherence.

Accordingly, differences between completers and dropouts were tested in bivariate analyses; age, gender, and marital status were significant in the CAU group. Women, younger individuals, and people who were single had a more difficult time remaining in the study than men, older individuals, and partnered patients. When testing this with all groups combined and statistically controlling for these group differences, only marital status was significant. However, this was merely the case when the psychological factors were included as well. It turned out that patients who indicated more depressiveness, more self-efficacy, and more social support were also more likely to drop out. This indicates that reducing symptoms in patients who are in greater need of treatment is difficult.

Furthermore, symptom reduction requires corresponding communication that builds tolerance against perceiving instant improvements but rather investing in active participation in treatment. This relates to the findings regarding improvement expectations: the more improvements patients expect due to the aftercare treatment, the more likely they are to also remain in the study after controlling for the aforementioned factors. Thus, working on those expectations right at the beginning of the therapy is imperative, especially if patients report high levels of depressiveness.

At the same time, it should be kept in mind that all the study participants had finished an intensive inpatient rehabilitation treatment and those still suffering from depression might have had a chronic or therapy-resistant medical condition. This might have led to the failure of aftercare and resulted in dropouts, which calls even more for individualized treatments addressing chronic depressiveness and the risks of relapse.

In contrast, the finding that patients with more self-efficacy and higher social support were more likely to drop out is noteworthy; this could be interpreted as unmet needs and expectations. At the same time, patients with high self-efficacy and social support might feel more capable to find better support or treatment. On the other hand, those with low self-efficacy and low social support seem to be a good fit for the online and F2F therapies provided in this study. More attention should be paid to patients with low expectations who appear difficult to cater to.

Regarding satisfaction with the aftercare, the participants in the online aftercare groups indicated higher satisfaction values than the participants in the F2F group. This can be explained by the advantage of requiring less time for commuting to the therapist [8] and the absence of the fear of stigmatization in online therapies [15]. Further studies are required for a precise clarification.

The fact that participants in the online aftercare groups rated the therapeutic relationship better than the participants in the F2F aftercare group might be due to the so-called online disinhibition effect, implying that people are more open to share their emotions and conflicts in a virtual space [35].

Limitations

This study was designed to investigate online therapies, compared with F2F therapy and CAU, in terms of symptom improvement. One main limitation was that the CAU group was not assessed regarding their expectations, due to practical and ethical considerations. Another limitation was the fact that dropouts were not specifically re-assessed because this was not planned within the study protocol [9].

However, further prospective and randomized studies are necessary to investigate the actual acceptance of online therapy opportunities and the prevention of dropout from (online) therapies and measurements [3-8,34,35]. Additionally, testing tailoring of the programs to the expectations and resources of the patients, specifically with regard to dropout and nonadherence, could provide additional insight. In the future, clinical trial registration should be prospective instead of retrospective.

Another noteworthy limitation is that 6023 patients were recruited but only 300 patients (about 5%) took part in the study. While reasons may vary from local and individual factors, it may also be the case that the program was interesting and fitting only for a very small subgroup of the addressed population.

In the future, the program should be designed in a way so that it better matches a larger proportion of the sample. Ideally, co-creative or co-design strategies that involve the target group could help, although this is typically very time and resource consuming.

Comparison With Prior Work

In this study, more patients dropped out than in other online intervention studies [6,13], which might be related with the longer follow-up period in this study. However, in their systematic review, Brown et al [11] did not find the intended duration of the program to be significant. More work is needed on the dose-response, along with testing whether the right length and intensity of therapy are related with lower dropout rates.

In the study by Zarski et al [3], 14% of the variance in treatment adherence could be explained by the variables of the HAPA model [22]. In our study, sociodemographic variables explained 4%-5% of questionnaire completion rates, which increased to 17%-20% when including social-cognitive variables related to the HAPA and additional health-related characteristics. While this percentage of the variance may appear small, one has to bear in mind that the predictor variables were assessed at baseline and dropouts at 15 or 18 months later, whereas in the study by Zarski et al [3], only a baseline measure and adherence 7 weeks later were analyzed.

In the systematic review by Brown et al [11], the duration of the interventions ranged between 3 and 20 weeks, with the majority lasting 8 weeks (n=25), 6 weeks (n=22), or 10 weeks (n=8). Thus, our study evaluated the intervention over a relatively long time frame, and more studies like this are required in the future to replicate our findings with larger samples.

Consistent with previous studies [6], dropout was less prevalent on a descriptive level in the group with more human feedback and less feedback filtered by the online delivery (ie, in the F2F group relative to all other groups). While previous studies [11] uncovered factors such as hardware or technical issues, this could also be assumed for this study, too, but few patients in the 2 online therapy groups actually reported this explicitly. On the other hand, the online therapy clearly overcame previously reported problems such as lack of time and work commitments [11,14], as well as commuting challenges to the physical intervention site [4,7,12,36].

Other problems such as disinterest and a diminishing desire to participate, perceptions that no further need for treatment would be required, feeling better after only a few modules, and perceiving the program as noneffective were found across all groups [11]. However, disappointment due to group assignment can be assumed, especially in the CAU group, and may be a reason for dropout and nonadherence [11].

The finding that poor health [6,11,17,20] could be related to poor adherence was also found in this study. However, this was only found with regard to depressiveness—the main symptom for assigning patients to psychosomatic rehabilitation treatment and its aftercare.

Other interrelations we revealed in our study also matched those in previous studies. For instance, higher education was (partially) related to lower dropout, probably because more self-reflection and eloquence make it easier to make use of the therapy [4,20,21]. No clear evidence regarding employment could be found in previous research [14], and our data supported this finding. While in previous studies, age was found to be related to the willingness to participate and remain in online research [14], we also found that younger patients were at a greater risk of dropping out [16].

Our finding that women were more likely to drop out from the study if randomized to the CAU group also matched previous findings (eg, [19]). If this could be attributed to the participating women being less technologically open and self-efficacious and less able to overcome technical problems [20], there would be a need for training and more adequate support. However, if this is related to family duties such as caring for children or other family members as found in previous studies [8,17,20], the online therapy might bring benefits both in terms of avoiding commuting times or eliminating the risk of leaving children or family members unattended at home. Nevertheless, this was not explicitly assessed in this study and calls for future research.

Online therapy might also bring the risk of multitasking at home and creating no clear detachment from family duties when spending time in therapy. Such patients may be just one wall away from family responsibilities, and this may also relate to difficulties explaining to family members that no disturbances are allowed. This could be addressed in terms of good planning, for instance, by having a babysitter both for attending a F2F treatment at a therapy site and, while ensuring confidentiality, when attending the online therapy at home.

Marital status and social support were revealed in our study as being significantly related to remaining in online interventions as has been shown in previous research [8,14,17]. However, an opposite pattern to previous studies [14,20] was found: Marital status was beneficial but more social support was not. Maybe a family member stepping in when problems, such as increasing family duties, made it more likely for individuals to remain in the therapy and the study, despite the difficulty. This underscores the importance of partnership or family for therapy adherence. On the contrary, dropout for those patients with high social support may be due to having perceived a mismatch in their expectations but then they got the help to find alternatives (while the partner or family individual is not able to do so). However,

these assumptions need to be researched further, in more detail, and systematically.

Matching previous findings, higher treatment expectancy regarding the treatment efficacy was related to greater study retention and adherence [14,20]. Remarkably, contrary to previous studies [3,14], in our study, self-efficacy was found to be directly related to dropout. As mentioned, this may be related to the study design and other factors relating to alternative treatment usage and self-help behavior. Thus, more work is needed to investigate this further.

Conclusions

This study showed that there are many different factors correlating with adherence to and dropout from online and F2F therapies. These variables should be addressed when allocating patients to their therapies and treating mental disorders.

Special focus should be given to women, younger patients, unpartnered patients, less educated patients, patients with more depressiveness symptoms, and those with fewer expectations. Tailored approaches should support these patients by meeting their needs and building optimistic expectations.

Acknowledgments

We thank Kureva Matuku for the literature search and Ronja Bellinghausen for editing a previous version of the manuscript. We appreciate the help by Kureva Matuku and Nerma Pasic for the language editing of the manuscript.

The Open Access Fee was provided by the Konfuzius Institut Bremen. The Konfuzius Institut Bremen had no role in the study design, study implementation, data collection, data analysis, manuscript preparation, or publication decision. The work is the responsibility of the authors.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the present study is a secondary analysis of previous randomized controlled trials. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
The differences between patients who completed questionnaires at T3 and those who dropped out from the study at T2 and T3.
[DOCX File , 15 KB - jmir_v23i11e31274_app1.docx]

Multimedia Appendix 2
Logistic regression models predicting completion of the questionnaire at T2 and T3 with only the patients receiving a therapy (ON1, F2F, ONL2).
[DOCX File , 15 KB - jmir_v23i11e31274_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 703 KB - jmir_v23i11e31274_app3.pdf]

References

1. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
2. 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 Feb 11;4(1):20 [FREE Full text] [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
3. Zarski A, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning good intentions into actions by using the Health Action Process Approach to predict adherence to internet-based depression prevention: secondary analysis of a randomized controlled trial. J Med Internet Res 2018 Jan 11;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]

4. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
5. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov 17;39(6):785-804 [FREE Full text] [doi: 10.1177/0145445515595198] [Medline: 26187164]
6. 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]
7. Karin E, Dear BF, Heller GZ, Crane MF, Titov N. "Wish you were here": examining characteristics, outcomes, and statistical solutions for missing cases in web-based psychotherapeutic trials. JMIR Ment Health 2018 Apr 19;5(2):e22 [FREE Full text] [doi: 10.2196/mental.8363] [Medline: 29674311]
8. Skea ZC, Newlands R, Gillies K. Exploring non-retention in clinical trials: a meta-ethnographic synthesis of studies reporting participant reasons for drop out. BMJ Open 2019 Jun 03;9(6):e021959 [FREE Full text] [doi: 10.1136/bmjopen-2018-021959] [Medline: 31164359]
9. Dahmen A, Gao L, Keller F, Lehr D, Becker P, Lippke S. Wirksamkeit webbasierter psychotherapeutische Nachsorge nach psychosomatischer Rehabilitation ? ein Test in zwei Randomized Controlled Trials. Die Rehabilitation (forthcoming) 2021:1.
10. Deutsche Rentenversicherung. URL: https://www.deutsche-rentenversicherung.de/DRV/DE/Experten/Infos-fuer-Reha-Einrichtungen/Grundlagen-und-Anforderungen/Konzepte-und-Positionspapiere/konzepte_positionspapiere.html [accessed 2021-10-01]
11. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
12. Buyruk Genç A, Amanvermez Y, Zeren SG, Erus SM. Early separations: Dropout from online and face-to-face counseling. Pegem Eğitim ve Öğretim Dergisi 2019 Sep 18;9(4):1001-1030 [FREE Full text] [doi: 10.14527/pegegog.2019.032]
13. Marks IM, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychol Med 2004 Jan 14;34(1):9-17. [doi: 10.1017/s003329170300878x] [Medline: 14971623]
14. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
15. Weightman M. Digital psychotherapy as an effective and timely treatment option for depression and anxiety disorders: Implications for rural and remote practice. J Int Med Res 2020 Jun;48(6):300060520928686 [FREE Full text] [doi: 10.1177/0300060520928686] [Medline: 32527170]
16. Karin E, Crane MF, Dear BF, Nielssen O, Heller GZ, Kayrouz R, et al. Predictors, outcomes, and statistical solutions of missing cases in web-based psychotherapy: methodological replication and elaboration study. JMIR Ment Health 2021 Feb 5;8(2):e22700 [FREE Full text] [doi: 10.2196/22700]
17. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
18. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: A meta-analysis. Computers & Education 2017 Feb;105:1-13 [FREE Full text] [doi: 10.1016/j.compedu.2016.11.003]
19. Retzer L, Reindl R, Zauter S, Richter K. Bevorzugen Frauen Face-to-Face-Beratung bei Insomnie? Somnologie 2021 Mar 02;25(2):151-154. [doi: 10.1007/s11818-020-00292-3]
20. Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. J Med Internet Res 2014 Oct 14;16(10):e232 [FREE Full text] [doi: 10.2196/jmir.3513] [Medline: 25316533]
21. Schmidt ID, Forand NR, Strunk DR. Predictors of dropout in internet-based cognitive behavioral therapy for depression. Cognit Ther Res 2019 Jun 16;43(3):620-630 [FREE Full text] [doi: 10.1007/s10608-018-9979-5] [Medline: 32879540]
22. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
23. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
24. Kobelt A, Nickel L, Grosch EV, Lamprecht F, Künsebeck HW. [Participation in psychosomatic outpatient care after in-patient rehabilitation]. Psychother Psychosom Med Psychol 2004 Feb;54(2):58-64. [doi: 10.1055/s-2003-812612] [Medline: 14872392]

25. Kobelt A, Grosch E. Indikation zur ambulanten Nachsorge (Curriculum Hannover) in der Psychosomatischen Rehabilitation. Psychotherapeut 2005 Sep;50(5):340-346. [doi: 10.1007/s00278-005-0438-x]
26. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
27. NCSS Statistical Software. URL: https://www.ncss.com/ [accessed 2021-10-01]
28. Rabung S, Harfst T, Kawski S, Koch U, Wittchen H, Schulz H. [Psychometric analysis of a short form of the "Hamburg Modules for the Assessment of Psychosocial Health" (HEALTH-49)]. Z Psychosom Med Psychother 2009;55(2):162-179. [doi: 10.13109/zptm.2009.55.2.162] [Medline: 19402020]
29. Mittag O, Meyer T, Glaser-Möller N, Matthis C, Raspe H. [Predicting gainful employment in a population sample of 4225 statutory pension insurance members covering a prognostic period of five years using a brief subjective prognostic employment scale (SPE Scale)]. Gesundheitswesen 2006 May;68(5):294-302. [doi: 10.1055/s-2006-926781] [Medline: 16773550]
30. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work Ability Index, 2nd revised edition. Helsinki, Finland: Finnish Institute of Occupational Health; 1998.
31. Schulte D. Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV). Zeitschrift für Klinische Psychologie und Psychotherapie 2005 Jun 1;34(3):176-187. [doi: 10.1026/1616-3443.34.3.176]
32. Nübling R, Kraft M, Henn J, Kriz D, Lutz W, Schmidt J, et al. [Testing the psychometric properties of the Helping Alliance Questionnaire (HAQ) in different health care settings]. Psychother Psychosom Med Psychol 2017 Nov 30;67(11):465-476. [doi: 10.1055/s-0043-111083] [Medline: 28854445]
33. Schmidt J, Nübling R. ZUF-8. Fragebogen zur Messung der Patientenzufriedenheit. In: Brähler E, Schumacher J, Strauß B, editors. Diagnostische Verfahren in der Psychotherapie. Göttingen, Germany: Hogrefe; 2002:392-396.
34. 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 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
35. Mehta VS, Parakh M. Web based interventions in psychiatry: an overview. Int J Ment Health Psychiatry 2015;01(03):1. [doi: 10.4172/2471-4372.1000108]
36. Lippke S, Dahmen A, Gao L, Guza E, Nigg CR. To what extent is internet activity predictive of psychological well-being? PRBM 2021 Feb;Volume 14:207-219. [doi: 10.2147/prbm.s274502]

Abbreviations

CAU: care as usual
F2F: face-to-face
HAPA: health action process approach
HAQ: Helping Alliance Questionnaire
HEALTH-49: Hamburg Modules for the Assessment of Psychosocial Health in Clinical Practice
ICD-10: International Classification of Disease, 10th revision
MANOVA: multiple analysis of variance
ONL1: Curriculum-Hannover-Online therapy in equivalence study
ONL2: Curriculum-Hannover-Online therapy in superiority study
PATHEV: Patient Questionnaire on Therapy Expectation and Evaluation
RCT: randomized controlled trial
T1: baseline
T2: 12-month follow-up measurement for the equivalence study; 9-month follow-up measurement for the superiority study
T3: 18-month follow-up measurement for the equivalence study; 15-month follow-up measurement for the superiority study.
WAI: Work Ability Index

Edited by G Eysenbach; submitted 15.06.21; peer-reviewed by L Beatty, D Erbe; comments to author 09.07.21; revised version received 26.08.21; accepted 13.09.21; published 03.11.21.

Please cite as:
Lippke S, Gao L, Keller FM, Becker P, Dahmen A
Adherence With Online Therapy vs Face-to-Face Therapy and With Online Therapy vs Care as Usual: Secondary Analysis of Two Randomized Controlled Trials
J Med Internet Res 2021;23(11):e31274
URL: https://www.jmir.org/2021/11/e31274
doi:10.2196/31274
PMID:34730541

Original Paper

Examining the Effectiveness of 3D Virtual Reality Training on Problem-solving, Self-efficacy, and Teamwork Among Inexperienced Volunteers Helping With Drug Use Prevention: Randomized Controlled Trial

Chih-Huei Chiang[1], MEd; Chiu-Mieh Huang[2], PhD; Jiunn-Jye Sheu[3], PhD; Jung-Yu Liao[4], PhD; Hsiao-Pei Hsu[5], MSN; Shih-Wen Wang[1], MEd; Jong-Long Guo[1], PhD

[1]Department of Health Promotion and Health Education, College of Education, National Taiwan Normal University, Taipei, Taiwan

[2]Institute of Clinical Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3]School of Population, Health University of Toledo, Toledo, OH, United States

[4]Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan

[5]Department of Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Jong-Long Guo, PhD
Department of Health Promotion and Health Education, College of Education
National Taiwan Normal University
No.162, Sec. 1, He-ping East Road
Taipei, 10610
Taiwan
Phone: 886 77493715
Email: jonglong@ntnu.edu.tw

Abstract

Background: Illegal drug usage among adolescents is a critical health problem. The Taiwanese government provides an accompanying volunteer program to prevent students who experiment with drugs from reusing them. An appropriate training program can improve volunteers' abilities to assist students using drugs. Problem-solving, self-efficacy, and teamwork are critical abilities for inexperienced volunteers who help with drug use prevention. By interacting with the animation or 3D virtual reality (VR) in the virtual scene, learners can immerse themselves in the virtual environment to learn, and 3D VR can increase learning opportunities and reduce the cost of human and material resources.

Objective: The aim of this study was to examine the effectiveness of spherical video-based virtual reality (SVVR) training in improving problem-solving, self-efficacy, and teamwork among volunteers who helped prevent adolescents from using illegal drugs.

Methods: This study used a randomized controlled design with a total of 68 participants in the experimental (n=35) and control (n=33) groups. The participants in the experimental group received the SVVR training program and their counterparts in the control group did not receive any training.

Results: Generalized estimating equation analyses indicated that the experimental group showed significant posttraining improvements in problem-solving and self-efficacy but not teamwork when compared with the control group.

Conclusions: The results of this study revealed that SVVR could improve participants' problem-solving skills and self-efficacy for assisting students in not using illegal drugs. However, future studies are suggested to develop effective SVVR to assist inexperienced volunteers in enhancing their teamwork abilities. We believed that introducing the training program to more sites can enhance volunteer training so that volunteers can have a better companionship effect when helping students quit drugs.

Trial Registration: ClinicalTrials.gov NCT05072431; https://clinicaltrials.gov/ct2/show/NCT05072431

(J Med Internet Res 2021;23(11):e29862) doi:10.2196/29862

KEYWORDS

3D virtual reality; volunteers; problem-solving; self-efficacy; teamwork

Introduction

Illegal drugs contain psychoactive substances and toxic chemicals that are harmful to adolescents. Numerous countries ban the use of illegal drugs by adolescents because it is highly associated with poor brain function [1], cognitive abilities [2], and academic performance [2]. In addition, drug use during adolescence predicts future drug-related disorders in adulthood [3]. Since recent years, new psychoactive substances are being widely used by adolescents in many countries, causing serious health consequences [4]. Thus, prevention of illegal drug and new psychoactive substance usage has become a critical and urgent issue for governments and schools worldwide.

The Taiwanese government is fully aware of the seriousness of adolescent drug use. In addition to setting up an intervention program to prevent students who experiment with drugs from reusing them [5], the Taiwanese government also provides an accompanying volunteer program (ie, the Sunshine volunteer program). To be a volunteer, an individual is expected to be enthusiastic and willing to be a life coach to adolescents with drug use history and assist them in quitting drugs. Before serving, volunteers are required to complete 12 hours each of basic and specific training on drug use prevention [6]. Currently, the 12-hour training program involves a lecture delivered using the traditional teaching method. Informed by a needs assessment before the study inception, volunteers told us that they would like to learn problem-solving skills and increase their self-efficacy to meet the job requirements when supporting students in not using illegal drugs. In addition, inexperienced volunteers would like to form good teams with students to work together in preventing drug use. Although the existing training program can increase professional knowledge on drug use prevention, it is insufficient in the aspects of problem-solving abilities, self-efficacy, and teamwork skills. Thus, it is critical to develop an innovative training program.

An appropriate training program can improve problem-solving abilities among volunteers [7]. Problem-solving pedagogy can assist learners to use existing knowledge as well as previous experiences and skills to generate effective and feasible solutions to present problems [8]. Therefore, the problem-solving teaching method is important to cultivate learners' abilities [9,10].

Self-efficacy is a critical construct of the social cognitive theory advocated by Bandura [11]. Self-efficacy refers to an individual's belief that he or she can perform the necessary behaviors to produce specific performance achievements. Self-efficacy reflects self-confidence in one's motivation, behavior, and social environment. These cognitive self-assessments influence all types of human experiences, including the goals that people strive for, energy they spend to achieve them, and possibility of reaching a certain level of behavioral performance [12]. The construct of self-efficacy has been widely employed in drug use prevention. For example, a previous study applying an electronic course program on drug use prevention indicated that the adolescent participants in the experimental group displayed better self-efficacy in terms of drug use resistance than their counterparts in the comparison group [5].

To help students stop using illegal drugs, teamwork is important for volunteers. Teamwork might include team building and collaborative discussions of the adolescents' daily life. A previous study indicated that teamwork might buffer the negative effects of the drinking environment on coworkers [13]. Moreover, a previous study indicated that teamwork is an essential component of a preventive program to help manage problem behaviors [14]. A proposed model also suggested that better teamwork among parenting partners might reduce an important set of family risk factors associated with drug use as well as behavioral and emotional problems in children [15].

Problem-solving, self-efficacy, and teamwork are critical abilities for inexperienced volunteers who help with drug use prevention. Lectures are the main components of the traditional training programs, which can be replaced by integrating with emerging technology. A virtual learning environment is provided by 3D virtual reality (3D VR). By interacting with the animation or 3D VR in the virtual scene, learners can immerse themselves in the virtual environment to learn, and 3D VR can increase learning opportunities and reduce the cost of human and material resources. Empirical research indicates that VR-based teaching can provide participants with new experiences and increase their learning ability [16].

The use of 3D VR with multiple perspectives and multisensory cues offers several potential benefits to education and training [17]. Furthermore, 3D VR promotes experiential and active learning, visualization and reification, learning in contexts impossible or difficult to experience in real life, motivation enhancement, collaboration fostering, adaptability, and evaluation and assessment. Previous studies also documented 3D VR as a useful intervention tool to train older adults [18,19]. Considering the success reported in the literature, this study projected the likelihood of the training effects on adult volunteers who helped prevent illegal drug use.

The 360-degree view Camera (VR Camera) is a useful hardware device to provide VR content. The VR-based 360-degree scenery video is a combination of image computing technology and VR wearables that allow learners to obtain an immersive experience. In recent years, VR techniques and spherical video-based virtual reality (SVVR) have made significant progress in numerous training programs, allowing learners to interact with the virtual world. For example, researchers applied this technology to train people with public speaking anxiety to improve their performance. The results showed that participants with moderate and high levels of speaking anxiety did show improvements after receiving training [20]. In addition, a previous study also showed that SVVR-based learning is effective for childbirth training in nursing education. These findings encouraged the use of SVVR-based training programs [21].

Researchers also applied VR technology in clinical settings and explored its effectiveness in the field of substance use. For example, VR technology and cue responses were used to explore smokers' cravings for smoking [22]. Saladin et al [23] found that 3D VR could induce physiological and psychological responses, and then produce the intention and behavior related to drug use. A previous study used VR to assist smokers in quitting smoking. It found that crushing cigarettes significantly

improved the score of nicotine addiction and the rate of quitting smoking, as well as reduced the loss of patients [24].

Another advantage of VR-based education is that such interactive training is more suitable for sensitive problems, such as drug use, to avoid ethical disputes. For example, it may not be appropriate to provide real drugs at the teaching site; however, virtual drugs can be allowed in a VR environment. To the best of our knowledge, this SVVR-based training is the first of its kind for volunteers. This study aims to develop an SVVR-based training program for volunteers helping in preventing illegal drug use by adolescents and evaluate the effectiveness of the training program.

Methods

Participants

Sunshine volunteers helped students with illegal drug use problem in schools. Each city and county had a volunteer group. After we sent an invitation letter to each of the 17 Sunshine volunteer groups, 5 of them agreed to participate. We invited all volunteers in the 5 groups (approximately 200) to participate in our study and randomized them into the experimental and control groups by assigning a number and flipping a coin. The selection criteria of participants were (1) helping at least 1 student with drug use problems, (2) having served for less than 2 years as a Sunshine volunteer, (3) having experience in using technology products, and (4) being able to operate SVVR. Recognizing service experience as a confounding factor in our sampling, we selected only the volunteers with less than 2 years of experience to minimize its effect. Training these novice volunteers was expected to produce greater improvements than enlisting experienced volunteers. There was no cybersickness incident that led to the exclusion of participants. [25]. After provision of their signed informed consent forms, 68 selected volunteers participated in the study and were allocated randomly into experimental and control groups by tossing a coin, leading to 35 and 33 participants in the experimental and control groups, respectively.

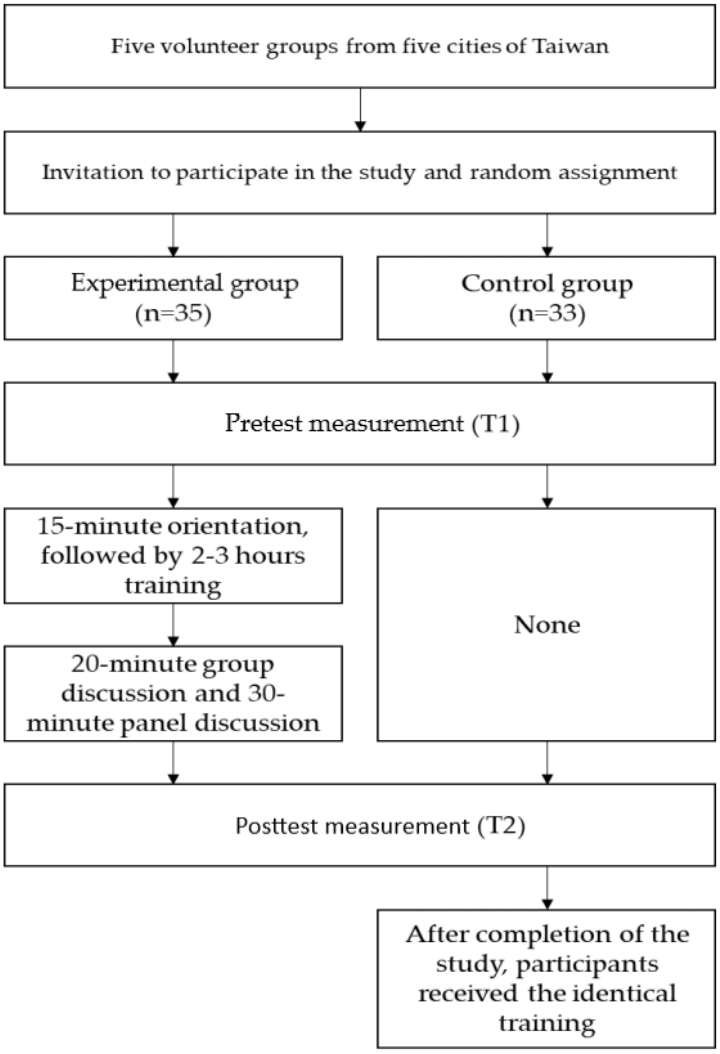
The participants in the control group received the SVVR intervention after completion of the study. The research team conducted one-on-one interviews to obtain the survey data. All participants successfully completed the program within approximately 3 hours.

According to Kirk [26], for an estimated effect size of 0.80, the approximate sample size required is 26 for each group when the power is set at 0.80 and type I error at 0.05. A previous 3D VR study yielded significant preintervention and postintervention improvements in psychological health with a similar sample size [19]. Another SVVR study with 32 participants in the experimental and control groups showed that providing childbirth education with SVVR resulted in superior learning performance [21]. Therefore, the sample size in this study (N=68) was large enough to detect training effects. Moreover, all participants finished the training program successfully.

Recruitment and Assessments

A flowchart outlining participant enrollment and assessments is presented in Figure 1. After selecting the five Sunshine volunteer groups, the research team approached the executive director and staff to explain the research purpose, method, and protocol. After obtaining permission to conduct the study, we posted recruitment messages and held meetings to invite potential participants who met the inclusion criteria so that the participants could fully understand the purpose of the study before providing their written consent. After the meeting, we conducted a 15-minute operation session with SVVR to evaluate the feasibility and acceptance of the program. Participants indicated their acceptance of this arrangement and reported that the SVVR program is very novel and different from their previous training experiences. The research team members collected their baseline data in a quiet room of a local school. During the implementation period, a 3D VR technical professional, staff of the volunteer group, and several trainers of the SVVR program were available to ensure the training proceeded smoothly. After the SVVR training, participants of the experimental group were involved in a 20-minute group discussion. After the group discussion, a counseling professional hosted a 30-minute panel discussion to answer the questions regarding how to effectively help adolescents quit substance use. The 20-minute group discussion and 30-minute panel discussion were conducted to assist participants to organize what they learned into 5 VR scenarios and use them during the future accompanying process with students. After the panel discussion, the participants provided written and verbal feedback to the research team. The participants were encouraged to contact the counseling professional in future to work together in supporting students to quit using illegal drugs. The control group did not receive the SVVR training at the same time but received it after completion of the study.

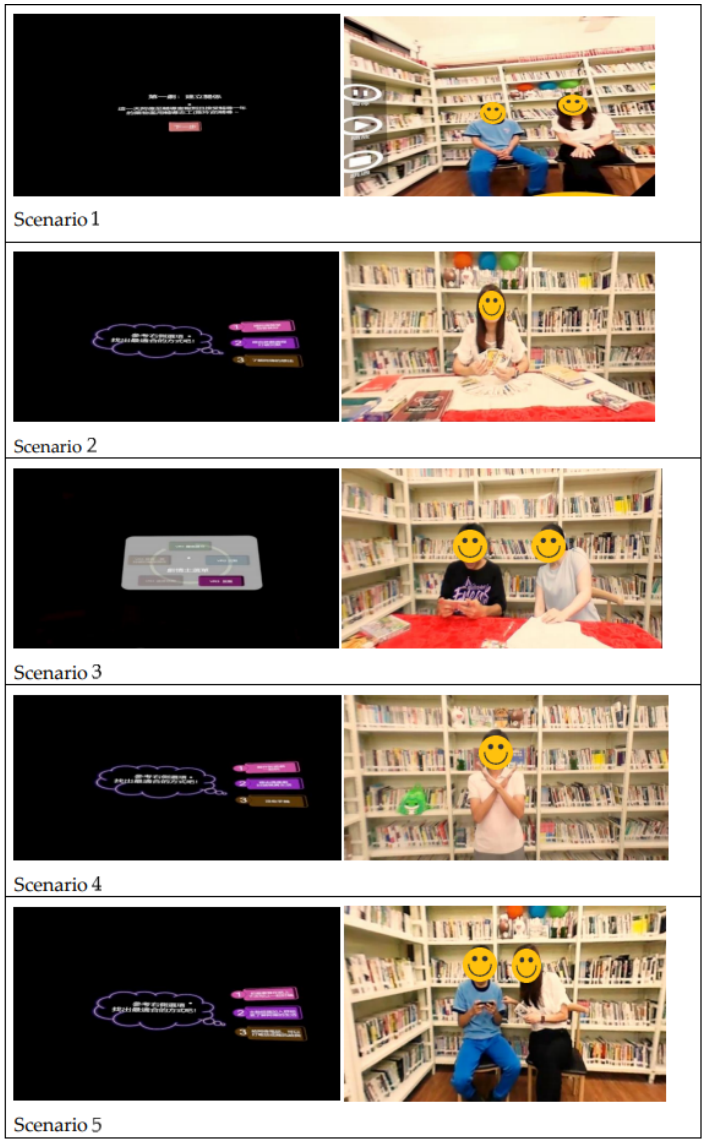
Figure 1. Flowchart showing the participant enrollment and assessment process.



Training Program Development

Before developing the SVVR program, the research team interviewed five volunteers face to face to understand how to help students quit substance use. Verbatim transcripts of the interviews were recorded. Consequently, the research team developed the SVVR program involving drug education, educational psychology and counseling, social work, nursing, and health education professionals to ensure the appropriateness of the SVVR program for the volunteers. The development of scenarios adopted a narrative approach in which each scenario included elements such as the setting, character, and plot. In addition, the context was derived from previous interviews of volunteers. This presentation format could make the scenarios more complete and better meet the needs of the volunteers. After the first draft of the training program was completed, the research team invited the volunteers and professionals to examine the five scenarios and provide their comments. The five scenarios are shown in Figure 2.

Figure 2. Five scenarios of spherical video-based virtual reality.



Program Components

The training program offered 5 VR scenarios and invited participants to actively engage in a 20-minute group discussion and a 30-minute panel discussion. Table 1 summarizes the five scenarios. Each scenario has a corresponding activity led by a drug use prevention professional to guide participants in having proper interactions with students. For example, the first scenario focused on ice breaking, greeting, mutual introduction, and welcome messages. In the second scenario, students chose three Tarot cards and interpreted the meanings of these cards. The third scenario was about how to face a situation when students play mobile games and neglect the participants. The fourth scenario was to use Kahoot, a game-based learning platform, to design 10 questions for exploring ways of facilitating communication between students and close peers using drugs. The fifth scenario taught the participants how to become friends or followers of the students on social media platforms (Line, Facebook, and Instagram) to assess the social groups and lifestyles of the students, regularly or irregularly post positive messages, offer students timely attention, and interact with the students when needed. The relationships of the training program components and outcome variables are presented in Table 1.

The research team conducted a 15-minute operation demonstration to introduce the purpose of the training before the SVVR training was formally launched. In this orientation, the 3D VR professional taught the participants how to wear the 3D VR helmets and select the VR scenes. The participants practiced multiple times to avoid dizziness.

Table 1. Activities of virtual reality training program and outcome variables.

Activities for VR[a] training program	VR1	VR2	VR3	VR4	VR5
Outcome variables					
Life skills					
Problem-solving (Information Children" [27]; Docktor and Heller [28])					
Step 1: Define the problem.	Warm up: Welcome, introduce each other, explain that you are a volunteer who is happy to accompany the students and ask them about their recent life.	Tarot card game: Pick up the first card and pretend the card reveals the students' current situation or problem. Participants interpret the card using information collected in advance.	N/A[b]	Help students identify if they have close peers (such as a buddy, boyfriend, or girlfriend) with drug use problems. When there is a binding relationship between the student and drug-using peers that prevents students from staying away, take the opportunity to advice students about the right way to interact with others.	N/A
Step 2: Gather information	Assist students in clarifying recent problems in life.	Tarot card game: Pick up the second card and encourage students to find the possible causes that resulted in the problems in their current lives.	N/A	N/A	N/A
Step 3: Generate possible solutions.	Express your concern, empathize, and invite students to brainstorm ideas about what has happened.	Tarot card game: Pick up the third card and suggest what should the student do in future to avoid similar problems.	N/A	N/A	N/A
Step 4: Evaluate possible methods and choose one.	N/A	N/A	Prepare yourself to handle students' problems and emotions.	N/A	Learn how to use one social media platform that is currently used by the students for timely communication with students.
Step 5: Implement the method to solve the problem and evaluate.	N/A	N/A	N/A	N/A	Learn how to send messages and emoticons or emojis using the app to help students solve their problems.
Self-efficacy empowerment					
Enactive self-mastery					
Coaching		Discuss with the students about what can be done to avoid problems in life.	N/A	N/A	N/A
Participation	Encourage students to continue this companionship.	N/A	N/A	N/A	N/A
Role modeling					

Activities for VR[a] training program	VR1	VR2	VR3	VR4	VR5
Demonstration	N/A	Participants also draw three Tarot cards to explain how to face difficult situations, such as asking for help from others, finding available resources, and calmly thinking about solving problems.	A VR professional shares his or her own experience and explains how to deal with problems effectively.	N/A	A VR professional demonstrates how to interact with students using social media platforms.
Mentoring	Introduce your role as the student's mentor and your willingness to assist him or her for at least 3 months or even until graduation.	N/A	N/A	Guide students to uncover one or more questions regarding close peers with drug use problems. Politely ask them to examine their interaction with their peers and tell them to think independently about "the good and bad" about people, affairs, and things.	A professional uses his or her mobile phone to demonstrate how to successfully interact with students for a long period of time.
Verbal persuasion					
Stimulation	N/A	N/A	Challenge participants to think about solutions.	N/A	N/A
Rewards	N/A	Indicate the benefits of facing problems and solving them.	N/A	N/A	N/A
Teamwork					
Committing to the development of teamwork	N/A	N/A	Tell participants to listen to students and respect their ideas so that students are willing to work together with the volunteers.	N/A	N/A
Performing assignments that elicit teamwork	N/A	N/A	N/A	N/A	Perform assignments with participants that elicit teamwork.
Focusing on the process	N/A	N/A	N/A	N/A	Focus on the process of observing students' social groups and lifestyles and provide positive feedback if the student makes any changes that reduce illegal drug use.
Providing meaningful feedback	N/A	N/A	N/A	Provide several useful suggestions to interact with close peers.	N/A

[a]VR: virtual reality.

[b]N/A: Not applicable.

Measurements

The sociodemographic variables of the participants assessed at baseline included their gender, age, the number of years completed as a Sunshine volunteer, marital status, employment, and the number of served students. The performance impact was assessed based on three variables including problem-solving skills, self-efficacy, and teamwork.

Problem-solving Skills

The problem-solving ability scale was modified from a previously evaluated instrument [29] comprising 6 items and was scored on a Likert-type scale ranging from 1 to 5, with higher scores indicating a higher level of problem-solving abilities in terms of drug use prevention.

Self-efficacy

The self-efficacy scale was a modified form of a previously tested scale [30] containing 6 items. Each item was scored on a Likert-type scale ranging from 1 to 5, with higher scores indicating a higher level of confidence in drug use prevention.

Teamwork

The team cooperation scale used was a modified version of a previously tested scale [31] with 4 items. Each item was scored on a Likert-type scale ranging from 1 to 5, with higher scores indicating a higher level of cooperation with team members. Analyses of the Cronbach α and exploratory factor analysis of each variable indicated that a single factor on each scale accounted for 64.26% to 65.20% of the variance, as shown in Table 2.

Table 2. Reliability and factor loadings of each variable.

Theme	Number of items	Cronbach α	Factor	Factor loading	Accounted variance (%)
Problem-solving skills	6	.888	1	0.764-0.842	64.64
Self-efficacy	6	.887	1	0.699-0.882	64.26
Teamwork	4	.811	1	0.589-0.899	65.20

Data Analysis

We conducted descriptive analysis of the demographic variables. In addition, two-tailed χ^2 tests were conducted to compare differences based on the gender, age, number of years completed as a Sunshine volunteer, marital status, employment, and number of served students. A generalized estimating equation (GEE) was used to explore the effects of problem-solving, self-efficacy, and teamwork. The GEE enables the understanding of the changes and their effects at the individual and group levels by estimating the average response of the population and is more advantageous than regression analysis that would enable prediction of the effect [32]. Statistical analyses were conducted using SPSS 22.0 (IBM Corp).

Ethics Statement

The study received approval from the Research Ethics Review Committee of National Taiwan Normal University (201805HS007).

Results

Demographic Data

Most of the participants in the experimental and control groups were female, aged 51 to 60 years and married. They had served as Sunshine volunteers for 7 or more years and were unemployed. The participants in the experimental and control groups had no statistically significant differences in terms of their demographic variables, as observed in Table 3. The participants in both groups were treated equivalently when comparing the intervention effects.

Table 3. Demographic data of participants (N=68).

Characteristic	Experimental group (n=35), n (%)	Control group (n=33), n (%)	χ^2 (*df*)	*P* value
Gender			1.019 (1)	.31
Male	10 (28.57)	6 (18.18)		
Female	25 (71.43)	27 (81.82)		
Age (years)			1.13 (2)	.57
≤50	7 (20)	7 (21.21)		
51-60	19 (54.29)	14 (42.42)		
≥61	9 (25.71)	12 (36.36)		
Marital status			0.41 (1)	.61
Married	34 (97.14)	31 (93.94)		
Others	1 (2.86)	2 (6.06)		
Employment history (years)			0.262 (1)	.61
≤6	13 (39.39)	11 (33.33)		
≥7	20 (60.61)	22 (63.67)		
Employment			0.883 (1)	.35
Yes	11 (31.42)	14 (42.42)		
Others	24 (68.57)	19 (57.58)		
Number of served students			0.243 (1)	.62
≤1	18 (51.43)	15 (45.45)		
≥2	17 (48.57)	18 (54.55)		

Improvements in Training Variables

Group differences in the patterns of change over time are shown in Figure 3. GEE analyses indicated a significant group × time interaction for problem-solving (β=3.055; Wald χ^2_1=4.757; *P*=.03) and self-efficacy (β=6.135; Wald χ^2_1=19.033; *P*<.001) but not for teamwork (β=1.025; Wald χ^2_1=1.172; *P*=0.28), as indicated in Table 4.

Figure 3. Changes in the experimental and control groups from pretest to posttest in terms of (A) problem-solving skills, (B) self-efficacy, and (C) teamwork.

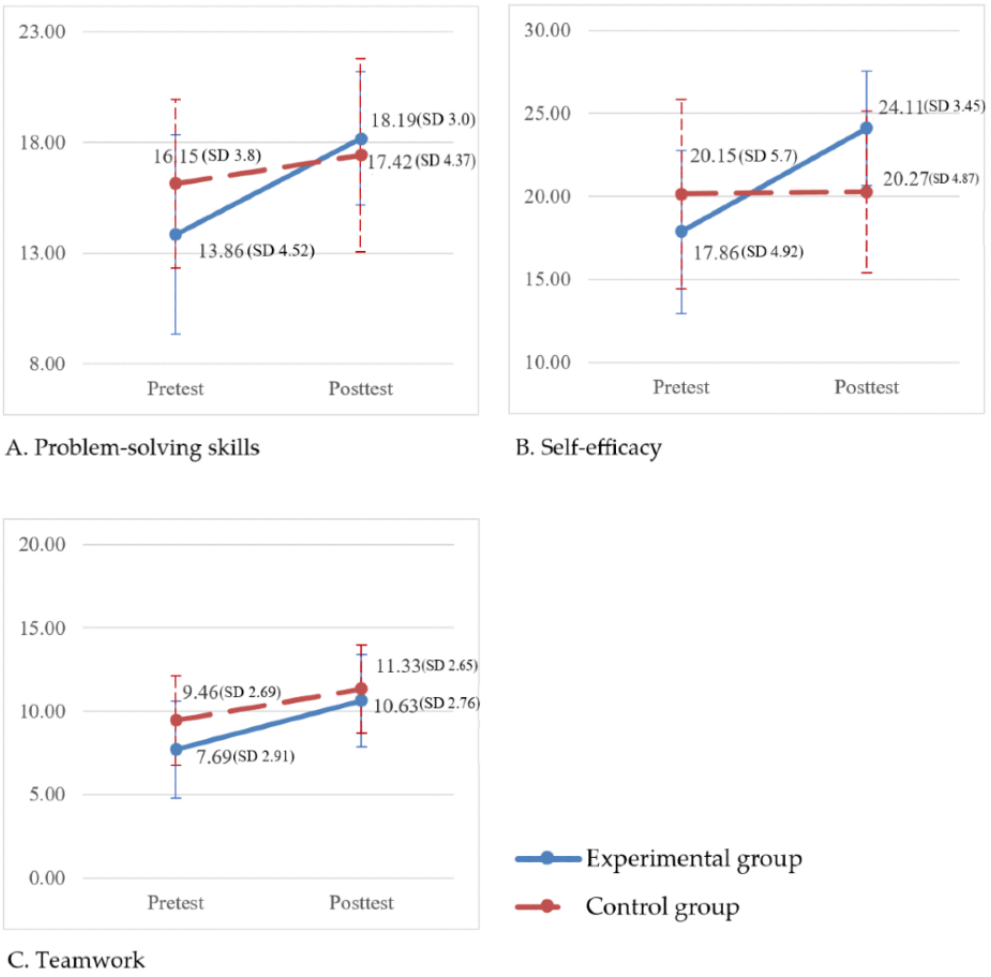


Table 4. Generalized estimating equation of outcome variables.

Life skill	Regression coefficient	SE	Wald χ^2 (*df*)	*P* value
Problem-solving				
Experimental group[a]	–2.283	1.107	4.250 (1)	*.04* [b]
Time (posttest)[c]	1.293	1.009	1.642 (1)	.20
Experimental group × time (posttest)[d]	3.056	1.401	4.757 (1)	*.03*
Self-efficacy				
Experimental group	–2.225	1.269	3.073 (1)	.08
Time (posttest)[c]	0.295	0.998	0.087 (1)	.77
Experimental group × time (posttest)	6.135	1.404	19.098 (1)	*<.001*
Teamwork				
Experimental group	–1.708	0.730	5.471 (1)	*.02*
Time (posttest)	1.880	0.468	16.112 (1)	*<.001*
Experimental group × time (posttest)	1.025	0.947	1.172 (1)	*.28*

[a]reference group: control group

[b]Italicized values indicate statistical significance at *P*>.05

[c]reference group: baseline

[d]reference group: control × baseline

Group and Panel Discussions

After the SVVR training, we invited the experimental group to discuss related responsive strategies. As shown in Table 5, the participants suggested responsive strategies including target setting, value identification, problem-solving skills, peer or parental influence, and environmental changes to assist students. During the panel discussion, the invited professionals and participants discussed responsive strategies including analyzing current situations, making multiple attempts, providing social support, and increasing social connections for students after quitting drugs considering their future life.

Table 5. Summary of responsive strategies for group and panel discussions.

Discussion	Responsive strategies
Group discussion: How do you help your students to avoid negative effects when drug-taking close peers still instill the wrong drug-taking concepts and ideas?	Target setting • Support students to understand their interests and expertise. There are goals to divert students' attention from drug use; the opportunities of success in discontinuing drug use are relatively high. (P1)a • By understanding the family background, friends, and working conditions of the students, we can provide responsive suggestions and tell them how and what they can achieve in future so that the students find the courage to live a new life. (P2) • Tell students that they will have more job opportunities if they possess a skill. (P1) • Help them find the focus of their life (things they are interested in) and gradually get rid of drugs. (P4) Help students find the type of work they are interested in and explain that all their efforts in life will help them achieve something. (P5) Value identification • Drugs harm health. For example, tell students that long-term use of ketamine will require them to use diapers for the rest of their life. (P9) Problem-solving • First understand the causes of drug use, analyze the advantages and disadvantages of drug use, discover the problems, and then help students solve the problems and return to a healthy life. (P3) • Inform the police about the drug supplier so that the students will not have access to drugs. (P12) Peer or parental influences • Let students know that they can leave peers who use drugs, meet with groups engaged in healthy activities, and make good new friends to support and encourage each other. (P4) • Cut off contacts with peers who use drugs. (P8) • Delete the phone numbers of peers who use drugs from the students' mobile phones. (P12) • Peers with risks of drug usage are also treated as persons needing care. (P12) • Parents are critical of students who use drugs. They should listen to and engage with these students. Environmental change • Choose the workplace carefully to avoid exposure to drugs. (P7) • Choose extracurricular activities and do not participate in activities that expose students to drugs. (P7) • Move out of the current living community to avoid exposure to drugs. (P8)
Panel discussion: Students ask what they can do if they quit drugs?	Analyzing current situations • First understand the current situations of the students' life and then provide specific suggestions. (P6) • First explore students' interests, talk about their needs, guide them to think about future career planning, and then guide them to focus on their dreams. If they work hard, they can achieve all the good things they want. (P10) Multiple attempts • Encourage students to participate in government-sponsored vocational training and explore more ways to establish their own career direction. (P1) • Work in an interesting workplace to see if students are suitable for the job position. (P2) • Forget the past and try to move forward. Keep the same job for at least 3 months before changing. (P2) • Lead the students to participate in academic and industry collaborations. (P5) • Help students find jobs. (P6) • Introduce them to appropriate work opportunities according to their personality and expertise. (P9) Giving social support • First address the students' emotions; let them take a break and let go of their blind spots; try to make them break through the bottleneck of life at this stage. (P5) • Support students till the end and try to help them. (P7) • Engage with students to find their interests; as long as they do not use drugs, everything is fine. (P8) • Support students to make sensible decisions and be their backup. (P9) • Get to know the students and help them. (P9) • Let students work or start a business together; provide consultation and assistance in this process. (P13) Increasing social connections • Encourage students to volunteer to care for the elderly and orphans and learn to help others and socialize. (P9)

[a]P: serial number of the participants.

Discussion

Principal Findings

To the best of our knowledge, this was the first study using SVVR in a volunteer training program. We adopted a randomized controlled design and recruited 68 participants from 5 cities to have enough power to detect the training effects. Although the score improvement aspects among the participants in the experimental group were superior to those shown by their counterparts in the control group in terms of problem-solving skills, self-efficacy, and teamwork after training, the GEE results revealed that the participants in the experimental group showed better improvements in problem-solving skills and self-efficacy compared with their counterparts in the control group but not in teamwork capabilities.

The SVVR training program improved the participants' problem-solving abilities when compared to their counterparts in the control group. Problem-solving is one of the life skills that was advocated by the World Health Organization [33]. Problem-solving skills involve identifying a problem, developing possible solution paths, and taking the appropriate course of action. Problem-solving skills can improve adolescents' problematic behaviors such as substance use [34]. The use of VR technology in the training program offers a novel opportunity for the development of problem-solving skills by providing learners with richer situations; this made the learning process more interesting and interactive and could improve the learners' motivation and attention, thus helping them explore new possibilities. A recent study supported the use of VR technology stating that it could increase users' interest and motivation, and potentially assist students in developing problem-solving skills [35]. Apart from the SVVR training program, our study conducted group and panel discussions. It contributed to the development of participants' problem-solving skills through discussion processes so that the volunteers could understand the importance of having discussions with the other participants.

Another important finding of the study is that the SVVR training program was effective in promoting volunteers' self-efficacy to help students quit drugs. Self-efficacy is one of the core concepts in Bandura's social learning theory, which refers to the degree of personal confidence in assessing whether a particular behavior is performed in a specific situation. It is a critical variable in predicting and interpreting future behavior [10]. A previous study pointed out that self-efficacy can stimulate behaviors and is the source of motivation [36]. The higher the self-efficacy, the higher is the number of expected healthy behaviors [37]. Self-efficacy could be a significant predictor for drug quitting among adolescents [5]. In addition, self-efficacy reflects the proficiency in personal abilities. An individual's self-efficacy can be improved through personal experiences of success [38]. Our findings advocate the use of SVVR for improving participants' self-efficacy, and they are consistent with the findings of previous studies [39,40]. A recent study indicated that using VR-based learning environments with student teachers helped students increase their self-efficacy and allowed them to be more innovative and creative [41]. Thus, using VR for training adolescents in self-efficacy offers high potential at present and in future.

A previous study identified significant differences between virtual and live simulations in terms of teamwork attitudes and communication skills in a randomized controlled trial that was conducted with 120 undergraduate medical and nursing students. These findings supported the potential of VR as a substitute for conventional team-based simulation training [42]. Numerous learning and performing tasks require teamwork. Team members may work concurrently and meet during some occasions. A recent review article summarizing trainings for health professionals indicated that most review studies evaluated the usability and acceptability of VR simulations whereas very few studies have measured the effects of VR simulations on the development of nontechnical skills such as teamwork [43]. Researchers indicated that although the importance of teamwork in health care is recognized, limited consensus exists regarding what it is, how it can most effectively be learned, and how it should be assessed [44]. We suggest that collaborative virtual environments should be developed in future training programs to allow participants to come together to complete a task. VR might have different educational or training effects on teamwork, suggesting that more research is needed in future to clarify its effectiveness.

Conclusions

We conclude that it is feasible to apply 3D VR technology in training volunteers who help with substance use prevention. The results of this study revealed that SVVR can improve participants' problem-solving and self-efficacy for assisting students in quitting illegal drug use. However, future studies are suggested to develop effective SVVRs to assist inexperienced volunteers in enhancing their teamwork abilities. We believed that it is promising to introduce this training program at more sites to enhance volunteer training so that volunteers can be better companions and help students quit drugs.

Acknowledgments

This research was funded by Ministry of Science and Technology, Taiwan (grant 106-2511-S-003-021). This article was subsidized by the National Taiwan Normal University, Taiwan, Republic of China.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1086 KB - jmir_v23i11e29862_app1.pdf]

References

1. Silveri MM, Dager AD, Cohen-Gilbert JE, Sneider JT. Neurobiological signatures associated with alcohol and drug use in the human adolescent brain. Neurosci Biobehav Rev 2016 Nov;70:244-259 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.06.042] [Medline: 27377691]
2. Camchong J, Lim KO, Kumra S. Adverse effects of cannabis on adolescent brain development: a longitudinal study. Cereb Cortex 2017 Mar;27(3):1922-1930 [FREE Full text] [doi: 10.1093/cercor/bhw015] [Medline: 26912785]
3. Shanks RA, Ross JM, Doyle HH, Helton AK, Picou BN, Schulz J, et al. Adolescent exposure to cocaine, amphetamine, and methylphenidate cross-sensitizes adults to methamphetamine with drug- and sex-specific effects. Behav Brain Res 2015 Mar;281:116-124. [doi: 10.1016/j.bbr.2014.12.002] [Medline: 25496784]
4. Helander A, Bäckberg M. New psychoactive substances (NPS) - the Hydra monster of recreational drugs. Clin Toxicol (Phila) 2017 Jan;55(1):1-3. [doi: 10.1080/15563650.2016.1217003] [Medline: 27549399]
5. Chang C, Liao J, Huang C, Hsu H, Chen C, Guo J. Evaluation of the effects of a designated program on illegal drug cessation among adolescents who experiment with drugs. Subst Abuse Treat Prev Policy 2018 Jan;13(1):2-12 [FREE Full text] [doi: 10.1186/s13011-017-0139-9] [Medline: 29338751]
6. Key points for the implementation of Sunshine volunteer service of the Ministry of Education. Ministry of Education. 2018. URL: https://edu.law.moe.gov.tw/LawContent.aspx?id=GL000333 [accessed 2020-12-24]
7. Merchant RM, Leigh JE, Lurie N. Health care volunteers and disaster response — first, be prepared. N Engl J Med 2010 Mar;362(10):872-873. [doi: 10.1056/nejmp1001737]
8. Häkkinen P, Järvelä S, Mäkitalo-Siegl K, Ahonen A, Näykki P, Valtonen T. Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills. Teachers and Teaching 2016 Jul;23(1):25-41. [doi: 10.1080/13540602.2016.1203772]
9. Proulx J, Maheux JF. From problem solving to problem posing, and from strategies to laying down a path in solving: taking Varela's ideas to mathematics education research. Constr Found 2017;13(1):160-167 [FREE Full text]
10. van Velzen JH. Complex problem solving in L1 education: senior high school students' knowledge of the language problem-solving process. J Educ Res 2016 Sep;110(6):634-641. [doi: 10.1080/00220671.2016.1171198]
11. Bandura A, Cliffs E. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Prentice Hall; 1986.
12. American Psychological Association. Teaching tip sheet: self-efficacy. 2009. URL: https://www.apa.org/pi/aids/resources/education/self-efficacy [accessed 2020-11-02]
13. Bennett JB, Lehman WE. Workplace drinking climate, stress, and problem indicators: assessing the influence of teamwork (group cohesion). J Stud Alcohol 1998 Sep;59(5):608-618. [doi: 10.15288/jsa.1998.59.608]
14. Graham K, Osgood DW, Zibrowski E, Purcell J, Gliksman L, Leonard K, et al. The effect of the Safer Bars programme on physical aggression in bars: results of a randomized controlled trial. Drug Alcohol Rev 2004 Mar;23(1):31-41. [doi: 10.1080/09595230410001645538] [Medline: 14965885]
15. Sanders MR. Community-based parenting and family support interventions and the prevention of drug abuse. Addict Behav 2000 Nov;25(6):929-942. [doi: 10.1016/s0306-4603(00)00128-3]
16. Mikropoulos TA, Natsis A. Educational virtual environments: a ten-year review of empirical research (1999–2009). Comput Educ 2011 Apr;56(3):769-780. [doi: 10.1016/j.compedu.2010.10.020]
17. Mantovani F. VR learning: potential challenges for the use of 3D environments in education training. In: Riva G, Galimberti C, editors. Towards Cyberpschychology: Mind, Cognition and Society in the Internet Age. Amsterdam: IOS Press; 2001:207.
18. Lin T, Huang C, Hsu H, Liao J, Cheng VY, Wang S, et al. Effects of a combination of three-dimensional virtual reality and hands-on horticultural therapy on institutionalized older adults' physical and mental health: quasi-experimental design. J Med Internet Res 2020 Nov;22(11):e19002 [FREE Full text] [doi: 10.2196/19002] [Medline: 33135666]
19. Cheng VY, Huang C, Liao J, Hsu H, Wang S, Huang S, et al. Combination of 3-dimensional virtual reality and hands-on aromatherapy in improving institutionalized older adults' psychological health: quasi-experimental study. J Med Internet Res 2020 Jul;22(7):e17096 [FREE Full text] [doi: 10.2196/17096] [Medline: 32706660]
20. Stupar-Rutenfrans S, Ketelaars LEH, van Gisbergen MS. Beat the fear of public speaking: mobile 360° video virtual reality exposure training in home environment reduces public speaking anxiety. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):624-633. [doi: 10.1089/cyber.2017.0174] [Medline: 29039704]
21. Chang C, Sung H, Guo J, Chang B, Kuo F. Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. Interact Learn Environ 2019 Sep:1-17. [doi: 10.1080/10494820.2019.1661854]
22. Bordnick PS, Graap KM, Copp H, Brooks J, Ferrer M, Logue B. Utilizing virtual reality to standardize nicotine craving research: a pilot study. Addict Behav 2004 Dec;29(9):1889-1894. [doi: 10.1016/j.addbeh.2004.06.008] [Medline: 15530734]

23. Saladin ME, Brady KT, Graap K, Rothbaum BO. A preliminary report on the use of virtual reality technology to elicit craving and cue reactivity in cocaine dependent individuals. Addict Behav 2006 Oct;31(10):1881-1894. [doi: 10.1016/j.addbeh.2006.01.004] [Medline: 16516397]
24. Girard B, Turcotte V, Bouchard S, Girard B. Crushing virtual cigarettes reduces tobacco addiction and treatment discontinuation. Cyberpsychol Behav 2009 Oct;12(5):477-483. [doi: 10.1089/cpb.2009.0118] [Medline: 19817561]
25. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. J Hum Reprod Sci 2011 Jan;4(1):8-11 [FREE Full text] [doi: 10.4103/0974-1208.82352] [Medline: 21772732]
26. Kirk RE. Statistics: An Introduction. Boston: Cengage Learning; 2005.
27. 5-steps to problem solving. Information Children. 2012. URL: http://www.informationchildren.com/5-steps-to-problem-solving/ [accessed 2020-10-16]
28. Docktor J, Heller K. Robust assessment instrument for student problem solving. In: Proceedings of the NARST 2009 Annual Meeting. 2009 Presented at: NARST 2009 Annual Meeting; April 17-20, 2009; Garden Grove, California p. 1-19 URL: https://groups.physics.umn.edu/physed/Talks/Docktor_NARST09_paper.pdf
29. Teamwork value rubric. Association of American Colleges and Universities. 2009. URL: https://www.aacu.org/sites/default/files/files/VALUE/Teamwork.pdf [accessed 2020-12-22]
30. Schwarzer R, Bäßler J, Kwiatek P, Schröder K, Zhang JX. The assessment of optimistic self-beliefs: comparison of the German, Spanish, and Chinese versions of the General Self-efficacy Scale. Appl Psychol 1997 Jan;46(1):69-88. [doi: 10.1111/j.1464-0597.1997.tb01096.x]
31. Hughes RL, Jones SK. Developing and assessing college student teamwork skills. New Directions for Institutional Research 2011 Apr;2011(149):53-64. [doi: 10.1002/ir.380]
32. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
33. Skills for health: skills-based health education including life skills: an important component of a child-friendly/health-promoting school. World Health Organization. 2003. URL: http://apps.who.int/iris/bitstream/handle/10665/42818/924159103X.pdf?sequence=1 [accessed 2020-11-02]
34. What does problem solving look like? Introduction to Problem Solving Skills. URL: https://ccmit.mit.edu/problem-solving/ [accessed 2020-11-02]
35. Araiza-Alba P, Keane T, Chen WS, Kaufman J. Immersive virtual reality as a tool to learn problem-solving skills. Comput Educ 2021;164:104121. [doi: 10.1016/j.compedu.2020.104121]
36. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WMP, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
37. Martinelli AM. Testing a model of avoiding environmental tobacco smoke in young adults. Image J Nurs Sch 1999;31(3):237-242. [doi: 10.1111/j.1547-5069.1999.tb00487.x] [Medline: 10528453]
38. Ormrod JE. Human Learning (7th edition). Harlow, England: Pearson; 2016.
39. Botella C, Baños R, Perpiñá C, Villa H, Alcañiz M, Rey A. Virtual reality treatment of claustrophobia: a case report. Behav Res Ther 1998 Feb;36(2):239-246. [doi: 10.1016/s0005-7967(97)10006-7]
40. Shu Y, Huang Y, Chang S, Chen M. Do virtual reality head-mounted displays make a difference? a comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. Virtual Reality 2018 Nov;23(4):437-446. [doi: 10.1007/s10055-018-0376-x]
41. Nissim Y, Weissblueth E. Virtual Reality (VR) as a source for self-efficacy in teacher training. Int Educ Stud 2017 Jul;10(8):52-59. [doi: 10.5539/ies.v10n8p52]
42. Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, 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;22(4):e17279 [FREE Full text] [doi: 10.2196/17279] [Medline: 32267235]
43. Bracq M, Michinov E, Jannin P. Virtual reality simulation in nontechnical skills training for healthcare professionals. Sim Healthcare 2019 Jan;14(3):188-194. [doi: 10.1097/sih.0000000000000347]
44. Weinstock P, Halamek LP. Teamwork during resuscitation. Pediatr Clin North Am 2008 Aug;55(4):1011-1024. [doi: 10.1016/j.pcl.2008.04.001] [Medline: 18675031]

Abbreviations

SVVR: spherical video-based virtual reality
VR: virtual reality

Edited by R Kukafka; submitted 06.05.21; peer-reviewed by LC Chang, CY Lin; comments to author 25.05.21; revised version received 28.07.21; accepted 12.08.21; published 02.11.21.

Please cite as:
Chiang CH, Huang CM, Sheu JJ, Liao JY, Hsu HP, Wang SW, Guo JL
Examining the Effectiveness of 3D Virtual Reality Training on Problem-solving, Self-efficacy, and Teamwork Among Inexperienced Volunteers Helping With Drug Use Prevention: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e29862
URL: https://www.jmir.org/2021/11/e29862
doi:10.2196/29862
PMID:34726606

Review

Impact of Virtual Reality-Based Therapies on Cognition and Mental Health of Stroke Patients: Systematic Review and Meta-analysis

Qi Zhang[1*], MD; Yu Fu[2*], MPhil; Yanhui Lu[1], MD, PhD; Yating Zhang[1], MA; Qifang Huang[1], MA; Yajie Yang[1], MA; Ke Zhang[3], BA; Mingzi Li[1], MD, PhD

[1]School of Nursing, Peking University, Beijing, China

[2]Department of Computer Science, Aberystwyth University, Wales, United Kingdom

[3]Department of Psychology, Shanxi Datong University, Datong City, China

[*]these authors contributed equally

Corresponding Author:
Mingzi Li, MD, PhD
School of Nursing
Peking University
38 Xueyuan Road
Beijing, 100191
China
Phone: 86 010 82805230
Email: limingzi2000@163.com

Abstract

Background: Stroke remains one of the major chronic illnesses worldwide that health care organizations will need to address for the next several decades. Individuals poststroke are subject to levels of cognitive impairment and mental health problems. Virtual reality (VR)-based therapies are new technologies used for cognitive rehabilitation and the management of psychological outcomes.

Objective: This study performed a meta-analysis to evaluate the effects of VR-based therapies on cognitive function and mental health in patients with stroke.

Methods: A comprehensive database search was performed using PubMed, MEDLINE (Ovid), Embase, Cochrane Library, and APA PsycINFO databases for randomized controlled trials (RCTs) that studied the effects of VR on patients with stroke. We included trials published up to April 15, 2021, that fulfilled our inclusion and exclusion criteria. The literature was screened, data were extracted, and the methodological quality of the included trials was assessed. Meta-analysis was performed using RevMan 5.3 software.

Results: A total of 894 patients from 23 RCTs were included in our meta-analysis. Compared to traditional rehabilitation therapies, the executive function (standard mean difference [SMD]=0.88, 95% confidence interval [CI]=0.06-1.70, *P*=.03), memory (SMD=1.44, 95% CI=0.21-2.68, *P*=.02), and visuospatial function (SMD=0.78, 95% CI=0.23-1.33, *P*=.006) significantly improved among patients after VR intervention. However, there were no significant differences observed in global cognitive function, attention, verbal fluency, depression, and the quality of life (QoL).

Conclusions: The findings of our meta-analysis showed that VR-based therapies are efficacious in improving executive function, memory, and visuospatial function in patients with stroke. For global cognitive function, attention, verbal fluency, depression, and the QoL, further research is required.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021252788; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=252788

(J Med Internet Res 2021;23(11):e31007) doi:10.2196/31007

KEYWORDS

virtual reality; stroke; cognition; depression; mental health

Introduction

Stroke is the second-highest cause of death worldwide and a leading cause of disability, contributing to approximately 3 million cases every year in China [1]. Survival rates of stroke have improved steadily over the past 2 decades; however, a longer survival length is often accompanied by the sequelae of long-term effects [2,3]. One such long-term effect of stroke is cognitive impairment [4]. It is estimated that approximately 80% of stroke survivors may experience new onset or worsening of cognitive impairment during their recovery [2,3]. Brain changes can affect 1 or more cognitive domains, including spatial awareness, praxis, perception, memory, language, and executive function [5]. Stroke-related cognitive deficits may interfere with functional recovery, the ability to (re-)acquire motor skills, and compromise independence [4,6,7], potentially exerting considerable influence on rehabilitation outcomes. Although stroke occurs as an acute event, it is a chronic condition that necessitates multidimensional and overwhelming treatment [8]. The neurological trauma of stroke survivors is irreversible and devastating; hence, after a stroke, patients face various stressors, which may trigger different aspects of their health, especially their mental health [9]. Mental health is defined as a state of complete happiness, which refers to our ability to enjoy life and cope with challenges [10]. Depression, anxiety, and stress are particularly common and persistent psychological problems following a stroke, with a high risk of relapse, even after a long period of remission [11]. Approximately 22%-40% of stroke survivors are affected by depression symptoms, 9.4%-36.7% by an anxiety disorder, and around 31% by poststroke stress [12]. Multiple studies have also shown that due to unsatisfactory sequelae and irreversible prognosis, patients with stroke always experience higher levels of emotional distress and ineffective coping than the general population, which poses adverse effects on the rehabilitation process and indirectly affects the patients' quality of life (QoL) [13,14].

Given the cognitive and psychological impacts on stroke survivors, strategies to support these individuals are considered a priority. Targeted rehabilitation can help address the clinically important cognitive and psychological consequences of stroke. However, the effect sizes of conventional rehabilitation are moderate at best, with high dropout rates, as patients, especially those with cognitive and psychological impairments, use rehabilitation services less efficiently and show less adherence to required changes in their lifestyle [15,16]. With underscored inadequacies of conventional rehabilitation, new high-tech innovations that use virtual reality (VR) are considered a potential avenue toward effective rehabilitation and may offer a supplementary platform to enhance the cognitive and psychological benefits after stroke.

VR is defined as a user–computer interface created with computer hardware and software, enabling the user to simulate interactions with environments that appear and feel similar to real-world objects and events through multiple sensory canals [17]. The term "virtual reality" was first coined in the late 1980s [18]. Until the late 1990s [19], VR-based rehabilitation was more widely used and sophisticated within health care systems as VR technology became more readily available and affordable [20,21]. Multiple recent studies have shown the positive role of VR in several clinical conditions, including acute pain [22], social anxiety disorder [23], and chronic obstructive pulmonary disease [21]. With the characteristics of immersion, interaction, and imagination, the application of VR in neurorehabilitation has also been growing rapidly in the recent years [24,25]. Currently, the neurorehabilitation applications of VR have been introduced in the field of neurorehabilitation for Parkinson's disease, Alzheimer's disease, brain injury, cerebral palsy, unilateral spatial neglect, and, especially, stroke [26-28]. Management of VR-based neurorehabilitation for stroke is highly related to recovery, reorganization, and neuroplasticity. VR can also exploit a brain mechanism known as embodied simulation, encouraging the patients' motivation and participation by allowing physical and emotional interactions with the environment through the digital medium [29]. In particular, VR offers a high level of flexibility and control over therapeutic tasks by automatically recording and tracking the user's performance [30,31], enabling the user to perform intensive training according to their ability and keeping the experience of interaction with therapeutic tasks enjoyable and compelling [25,32].

Due to tremendous benefits for health and the relatively inexpensive medium for rehabilitation training, a significant amount of work focusing on the effects of VR has been proposed for the rehabilitation of deficits following stroke [8,25]. Multiple recent randomized controlled trials (RCTs) on the effects of VR have supported the use of VR in stroke rehabilitation to improve cognitive and psychological outcomes [33-36]. In contrast to studies supporting VR-based rehabilitation, some researchers have argued that there are no or only minor effects detected in global cognitive function, memory, depression, or the QoL [34,37,38].

Although there are several systematic reviews conducted on the contribution of VR-based interventions to cognition rehabilitation, no unified conclusion has yet been reached. Anna et al [25] conducted a meta-analysis by summarizing RCTs published before June 2017 and supported VR as an adjunct to cognition rehabilitation. However, the number of included studies was only 2, suggesting a tenuous finding.

Another recent review [39] extensively searched relevant trials conducted before November 2019. It examined the effectiveness of VR-based therapies on both global cognition and domain-specific cognition poststroke, suggesting that VR therapy is not superior to control interventions in improving both global cognition and domain-specific cognition in patients with stroke. However, their conclusions were also merely based on data from 2 studies, resulting in insufficient statistical power. A systematic review by Zhang et al [40], only focusing on global cognitive function, also supported no benefits of VR for cognition. Furthermore, the psychological benefits of VR have also received more attention recently. A systematic review [8] indicated that exercise-based VR interventions are potentially valuable as a support in improving psychological outcomes. Nonetheless, due to the diversity of study design and insufficient data, the authors only described their results and did not conduct a meta-analysis. Hence, the overall effect of VR-based therapies

on improving cognition and mental health in patients with stroke remains unclear.

The evidence is sparse, and a comprehensive picture of the effects of VR-based therapies is needed. With ongoing advancements in VR, a plethora of original studies focused on cognitive recovery and mental health after stroke has been recently published. This makes it possible to review VR applications and guide future design and implementation of VR technology in clinical practice. Thus, our review aimed to comprehensively examine the effects of VR-based therapies on cognition, the QoL, and depressive symptoms in patients with stroke.

Methods

This systematic review was registered with Prospero, the International Prospective Registry of Systematic Reviews (registration no. CRD42021252788). We conducted this systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.

Literature Search Strategy

RCTs from PubMed, Embase, MEDLINE (Ovid), Cochrane Library, and APA PsycINFO databases were comprehensively searched. RCTs published in English before April 15, 2021, investigating the impact of VR-based therapies on the cognition and mental health of patients with stroke were included. In addition, we conducted reference tracking on the published trials and meta-analysis reviews in this field to make sure all relevant studies were included. The search strategies, including terms for stroke, VR, cognition function, the QoL, and depression, are presented in Multimedia Appendix 1.

Inclusion Criteria

Types of Trials

We only included RCTs that were peer reviewed and written in English. Reviews, single-case studies, dissertations, conference papers, and abstracts were excluded.

Types of Participants

Patients aged 18 years or older with stroke were included. Patients with stroke were identified by any available diagnostic criteria, such as brain computed tomography, magnetic resonance imaging, or other standards and consensus by clinicians. To maximize the number of meta-analyses, we did not restrict the search to any specific stroke population.

Types of Interventions

VR-based interventions include either single-component interventions or multiple-component interventions for patients with stroke. VR had to consist of a screen or a head-mounted device, including games with immersive, semi-immersive, and non-immersive systems, simulating virtual environments using computers, video consoles, mobile apps, and VR. The intervention setting, duration, and frequency were not restricted. Participants in the control group could undergo usual care or non-VR interventions.

Types of Outcome Measures

The primary outcomes were global cognition and domain-specific cognition (eg, attention, executive function, memory, psychomotor speed, verbal fluency). According to Isabelle Dor's [10] classification, the components of mental health include emotional well-being/QoL and psychological and social well-being. Our systematic review specifically included depression, stress/distress, anxiety, coping competence, overall mental health, the QoL, and self-efficacy.

Study Selection and Data Extraction

Two reviewers independently assessed studies obtained from the database searches in three phases: title, abstract, and full-paper screening. Data were extracted by the same two reviewers from trials using a Microsoft Excel spreadsheet. The information extracted from each trial included the first author, year, setting, sample size, participant ages, details of intervention and the control, outcomes, and corresponding measures. Disagreements were resolved through a discussion with a senior investigator.

Quality Appraisal

The Cochrane risk-of-bias tool was used to assess the quality of eligible trials, focusing on sequence generation, allocation concealment, blinding, attrition bias, completeness of outcome data, and other sources of bias. Any disagreements were settled by a discussion with a third person.

Data Synthesis and Analysis

The standard mean difference (SMD) and weighted mean difference (WMD) with a 95% confidence interval (CI) were determined for continuity data. Statistical heterogeneity was assessed using the Cochran Q test and the Higgins I^2 statistical test. The results implied low-level heterogeneity when $I^2 < 50\%$, and we used a fixed-effects model to fix the issue. The results implied moderate or high heterogeneity when $I^2 \geq 50\%$, and subgroups were determined based on the different characteristics of the studies to identify the sources of heterogeneity. Sensitivity analysis was performed by comparing the effect sizes and CIs of the remaining RCTs after removing each included RCT at a time. Descriptive analysis was used to deal with nonmergeable data. Potential publication bias was assessed with funnel plots and the Egger test if the group included 10 trials or more [41]. All meta-analyses were performed in RevMan 5.3 (Nordic Cochrane Center, Cochrane Collaboration, Copenhagen, Denmark) for outcomes that were evaluated in at least 2 of the included RCTs.

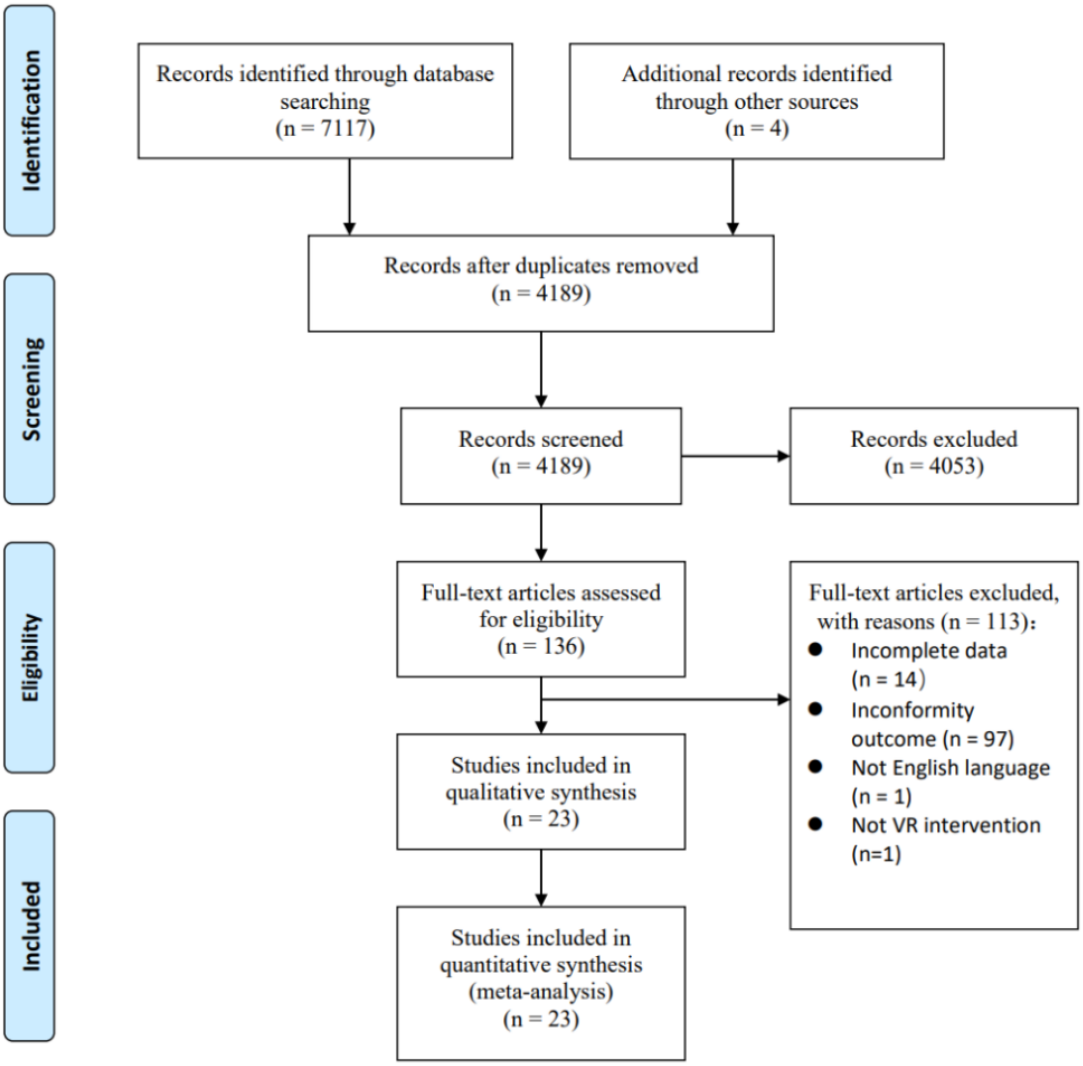
Results

Study Selection

A total of 7117 studies were identified from the 5 databases searched. A further 2932 studies were identified through manual searching, including searching for systematic reviews, and included studies' reference lists. After 261 duplicates were removed, 4189 full-text manuscripts were identified by screening their titles and abstracts. After full-text reviews, 23 studies (including 894 participants) from 8 countries satisfied

the inclusion criteria and were eventually included in our systematic review (Figure 1).

Figure 1. Flowchart of the study selection process. VR: virtual reality.



Study Characteristics

Multimedia Appendix 2 shows the study characteristics and patient demographic data of the included trials. Sample sizes ranged from 18 to 145, for a total of 894 patients with stroke. In total, 11 (47.8%) of the 23 trials were conducted in Korea [34,38,42-50], 3 (13%) in Portugal [51-53], 2 (8.7%) in Spain [54,55], 2 in China [36,37], 2 in Australia [56,57], and 1 (4.3%) each in Lithuania [58], Brazil [59] and Turkey [60].

Regarding the types of VR-based therapies, 16 (69.6%) trials administered VR as a singular session [36-38,44,45,47-49,51,52,54-57,59,60], 5 (21.7%) trials were categorized as having used VR-based therapies plus occupational therapy [43,46,50,53,58], and 2 trials as having used VR-based therapies plus computer-assisted cognitive rehabilitation [34,42]. The frequency and duration of VR exposure varied considerably between trials. In our review, the duration of VR-based therapies ranged from 3 to 10 weeks, with the majority being 4 weeks, and the frequency of intervention varied from 2 to 5 times per week. The control group underwent conventional rehabilitation therapy [36-38,44,47,48,51,54,56,57,59] or other support interventions such as occupational therapy [43,45,46,49,50,55,58], computer-assisted cognitive rehabilitation [34,42], and other rehabilitation therapies [52,53,60].

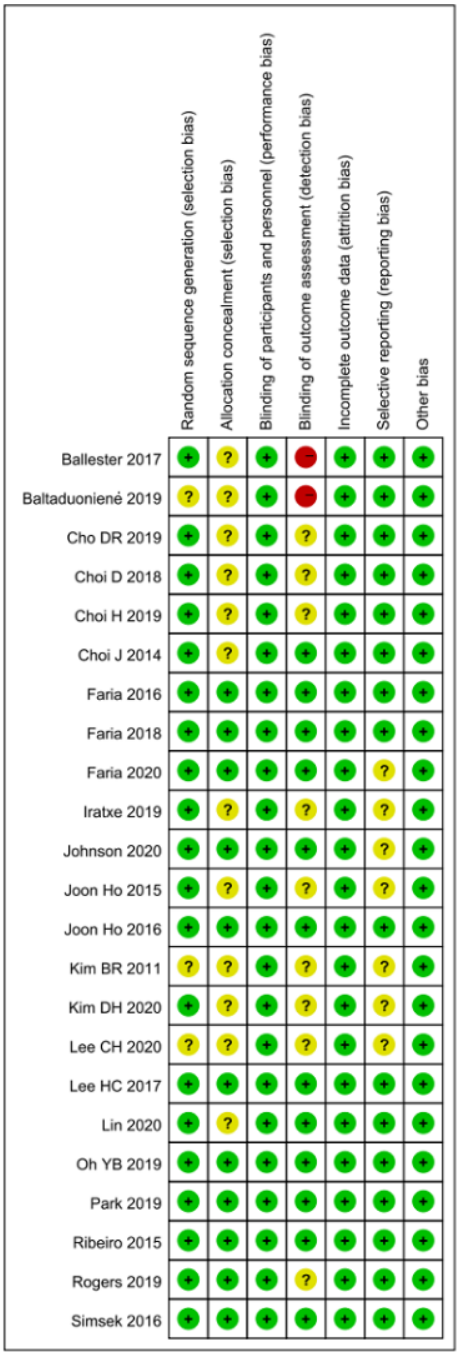
Of the 23 included trials, outcomes of our interest included global cognitive function (10 trials, 43.5%), executive function (5 trials, 21.7%), memory (5 trials), verbal fluency (2 trials, 8.7%), visuospatial ability (2 trials), attention (6 trials, 26.1%), depression symptom (5 trials), and the QoL (7 trials, 30.4%). Although the tools used for outcome evaluation varied across trials, all were valid scales, and the process of data collection was carried out by experienced sta . Tools used to evaluate global cognitive function included the Montreal Cognitive Assessment (MoCA) [52,53,56,58], the Mini-Mental State Examination (MMSE) [34,45,49,51], and the Loewenstein Occupational Therapy Cognitive Assessment (LOTCA) [38,42]. Five trials evaluated the executive function domain using the CogState Groton Maze Learning Task [56], the Tower of London Test [34], the Stroop test [48], Trail-Making Test-B [51], and the the Digit Span Test [52]. Tools used to evaluate memory of patients with stroke included the visual recognition test [42], the visual span test [34], Wechsler Memory Scale-III [52], the Digit Span Test [54], and the Trail-Making Test [48]. The Digit Span Test [34] and Wechsler Memory Scale-III [52]

were used to assess the verbal fluency of patients. Visuospatial ability was evaluated using the Motor-Free Visual Perception Test-3 [43] and the visual span test [34]. Attention was evaluated using the Single Letter Cancelation Test [53], Color of Color Word in Word–Color Test [34], and the Trail-Making Test [48,51,52,54]. The Hamilton Scale [36,46,55], the Beck Depression Inventory [37], and the Neurobehavioral Functioning Inventory [56] were used to assess the level of depression. The Stroke Impact Scale [38,47,50] was primarily used to assess the QoL, followed by the Short-Form 36 Health Survey Questionnaire [59], the Short-Form 8 Health Survey Questionnaire [44], the Nottingham Health Profile [60] and the EuroQoL Five Dimensions Questionnaire [57].

Risk-of-Bias Assessment

The risk of bias is summarized in Figure 2. In general, all 23 trials included in our review showed an acceptable risk of bias. The randomization sequence was adequately generated in 20 trials (86.9%), and 11 trials (47.8%) adequately concealed allocation. We categorized the risk of performance bias in all studies as low because blinding of participants and personnel was not possible in our systematic review. Approximately half of the trials blinded outcome assessors, and the risk of detection bias of these trials was judged as low. Of the 23 trials, 16 (69.6%) had no dropouts or used the intention-to-treat principle to compensate for dropouts; therefore, their risk of reporting bias was rated as low. All trials were categorized as having a low risk of bias in terms of attrition and other bias.

Figure 2. Risk-of-bias summary.



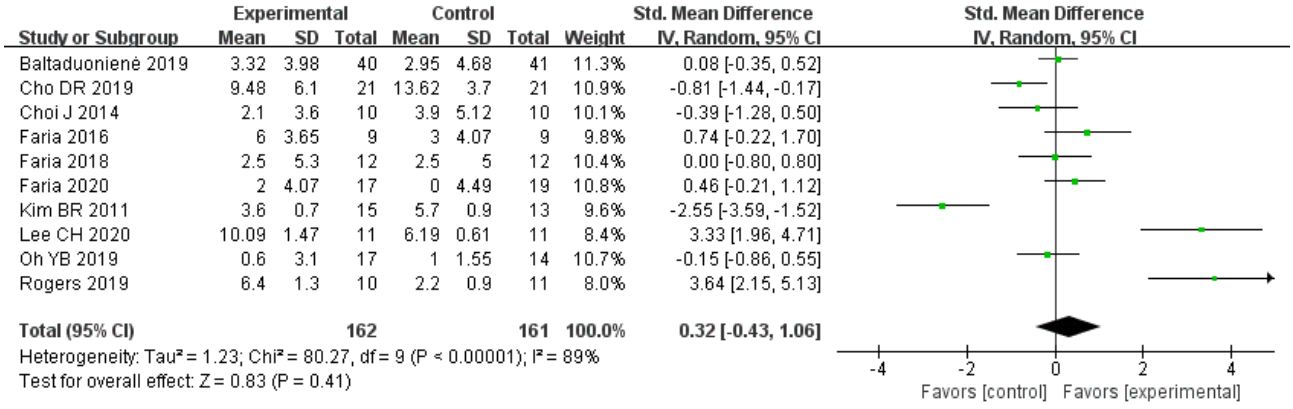
Results of the Meta-analysis

The outcomes of global cognitive function, domain-specific cognition, and mental health were evaluated using various tools in the included trials. Domain-specific cognition included executive function, memory, verbal fluency, visuospatial ability, and attention in our review. Depression and the QoL were synthesized in our review. The change scores from baseline to final values were used in our final efficacy analysis. The results of our analysis of each outcome are as follows.

Global Cognitive Function

As shown in Figure 3, effect sizes of global cognitive function could be generated for 10 trials [34,38,42,45,49,51-53,56,58] with 323 patients. A random-effects model was used as heterogeneity existed in our review (I^2=89%, P<.001). Compared to the control, there was no evidence that VR-based therapies can significantly improve global cognitive function for patients with stroke (SMD=0.32, 95% CI=–0.43-1.06, P=.41).

Figure 3. Forest plot for VR on global cognitive function. VR: virtual reality; CI: confidence interval.

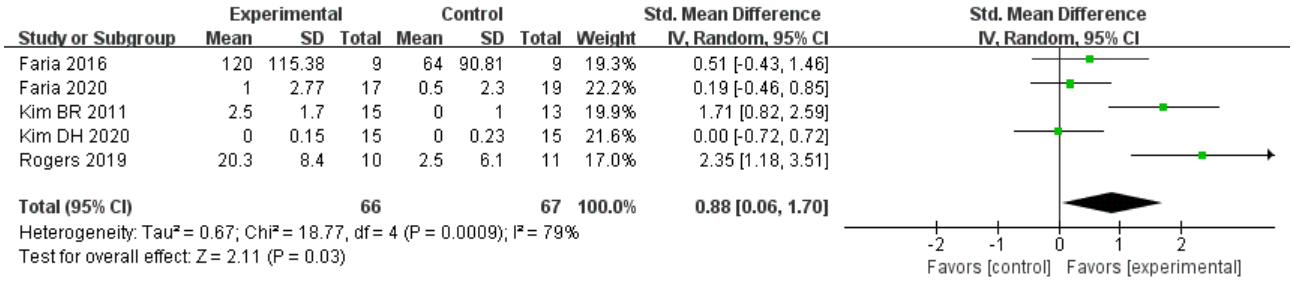


In the subgroup that included VR versus the usual-care control group, the effect was significant (SMD=2.51, 95% CI=0.52-4.5, P=0.01; I^2=87%, P<.001; high heterogeneity; N=61), while versus the other-support control group, no significant effect was detected (SMD=–0.41, 95% CI=–0.99-0.17, P=.16; I^2=79%, P<.001; high heterogeneity; N=262). The subgroup interaction was significant (χ2=7.65, P=.01, I^2=86.9%). In the subgroups of duration (less than 1-month or over 1-month intervention) and measurement instruments (MMSE, MoCA, or LOTCA), the subgroup interactions were nonsignificant (duration: χ2=0.12, P=.73, I^2=0%; measurement instruments: χ2=2.83, P=.24, I^2=29.2%), and heterogeneity was still significant in most of these subgroups.

Executive Function

The effect of VR-based therapies on executive function was measured in 5 trials [34,48,51,52,56] including 133 patients. The pooled results with a random-effects model showed that VR-based therapies can significantly improve executive function compared to the control (SMD=0.88, 95% CI=0.06-1.70, P=.03), with high heterogeneity (I^2=79%, P<.001; Figure 4). In the subgroups of the control groups (usual-care or other-support control) and duration (less than 1-month or over 1-month intervention), the subgroup interactions were nonsignificant (control: χ2=0.91, P=.34, I^2=0%; duration: χ2=1.46, P=.23, I^2=31.5%). Heterogeneity still existed in the subgroups.

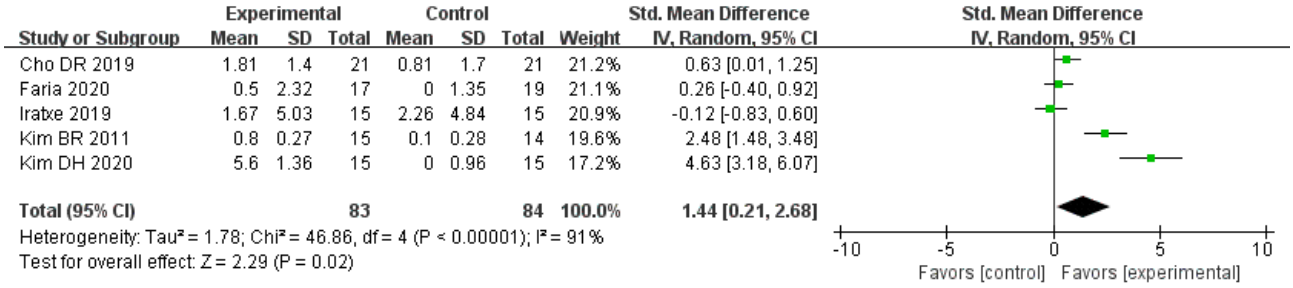
Figure 4. Forest plot for VR on executive function. VR: virtual reality; CI: confidence interval.



Memory

The effect of VR-based therapies on memory was measured in 5 trials [34,42,48,52,54] involving 167 participants. Based on a random-effects model, our results showed a beneficial effect of VR-based therapies on enhancing memory in patients with stroke (SMD=1.44, 95% CI=0.21-2.68, P=.02), with high heterogeneity (I^2=91%, P<.001; Figure 5). The control and duration subgroup interactions were nonsignificant (control: χ2=0.12, P=.72, I^2=0%; duration: χ2=0.88, P=.35, I^2=0%). Moderate heterogeneity existed in all subgroups.

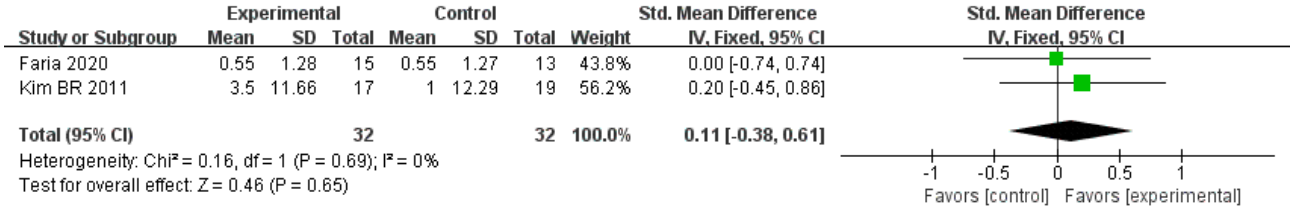
Figure 5. Forest plot for VR on memory. VR: virtual reality; CI: confidence interval.



Verbal Fluency

The effect of VR-based therapies on verbal fluency was measured in 2 trials [34,52] involving 64 patients. Using a fixed-effects model, our meta-analysis showed no statistical significance on verbal fluency (SMD=0.11, 95% CI=–0.38-0.61, *P*=.65), with no heterogeneity (I^2=0%, *P*=.65; Figure 6).

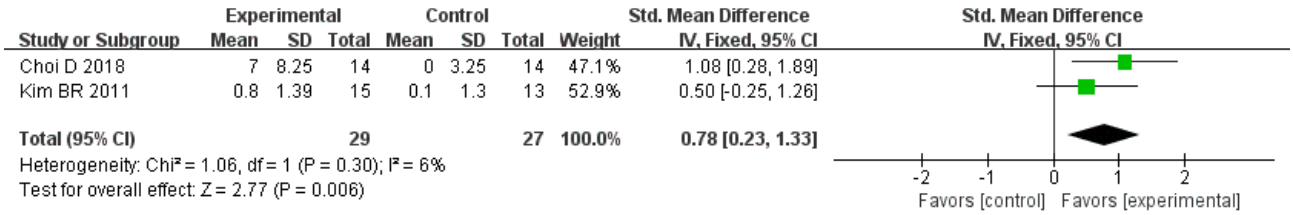
Figure 6. Forest plot for VR on verbal fluency. VR: virtual reality; CI: confidence interval.



Visuospatial Ability

The effect of VR-based therapies on visuospatial ability was measured in 2 trials [34,43] involving 56 patients. We used a fixed-effects model for pooling the results, and the overall result showed that VR-based therapies have a significant effect on visuospatial ability compared with the control (SMD=0.78, 95% CI=0.23-1.33, *P*=.006), with no heterogeneity (I^2=6%, *P*=.30; Figure 7).

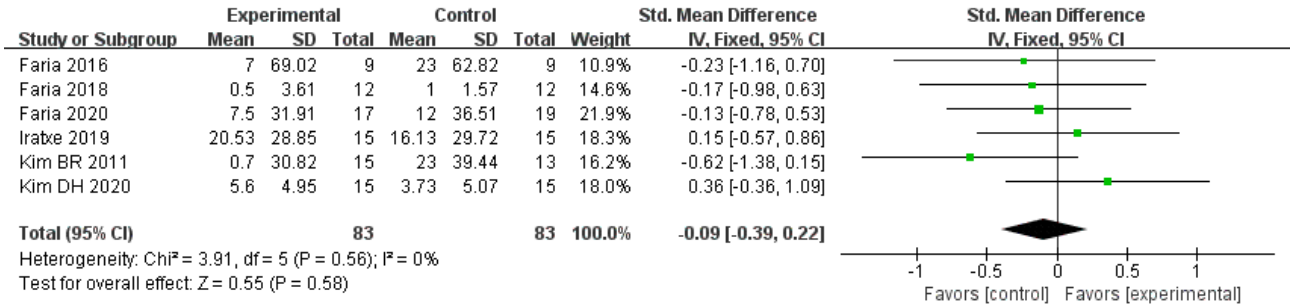
Figure 7. Forest plot for VR on visuospatial ability. VR: virtual reality; CI: confidence interval.



Attention

The effect of VR-based therapies on attention was measured in 6 trials [34,48,51-54] involving 166 patients. We used a fixed-effects model for pooling the results, and the overall result presented no significant effect on attention between groups (SMD=–0.09, 95% CI=–0.39-0.22, *P*=.58), with no heterogeneity (I^2=0%, *P*=.56; Figure 8).

Figure 8. Forest plot for VR on attention. VR: virtual reality; CI: confidence interval.



Depression

The effect of VR-based therapies on depression symptoms was measured in 5 trials [36,38,46,55,56] including 255 participants. Lower scores indicated better psychological states. Based on a random-effects model, our results indicated that there was no statistical significance between the VR and control groups (SMD=0.20, 95% CI=–0.25-0.64, *P*=.39), with moderate heterogeneity (I^2=56%, *P*=.06; Multimedia Appendix 3).

QoL

The effect of VR-based therapies on the QoL was measured in 7 trials [37,44,47,50,57,59,60] including 272 patients. Using a fixed-effects model, our results showed that VR-based therapies had no significant beneficial effect on the QoL (SMD=0.07, 95% CI=–0.17-0.31, *P*=.55), with low heterogeneity (I^2=12%, *P*=.34; Multimedia Appendix 3).

Publication Bias and Sensitivity Analyses

Funnel plots for global cognitive function showed symmetry. The Egger test of global cognitive function was not significant (*P*=.29, Multimedia Appendix 4), indicating no significant publication bias in our review. A sensitivity analysis was conducted, excluding individual trials one by one, to confirm the results of our meta-analysis. We found that except for the outcomes of executive function, the rest of the outcomes showed no substantial modification of the overall effect.

Discussion

Principal Findings

Our meta-analysis, based on 23 RCTs, indicated that VR therapy is an effective method of improving executive function, memory, and visuospatial ability on patients with stroke. However, current evidence fails to support the effects of VR-based therapies in improving global cognitive function, attention, verbal fluency, depression, and the QoL. Unlike prior systematic reviews [8,25] primarily focused more on either one type of VR, such as exercise-based VR, or one aspect of cognition indicators, such as only global cognitive function, our review [40] comprehensively evaluated the effect of VR-based therapies on global and domain-specific cognition and mental health outcomes.

VR-based therapies have demonstrated efficacy in the diagnosis, physical rehabilitation, and cognitive rehabilitation of individuals with neurocognitive disorders [17,19]. A growing body of studies has confirmed that the application of VR is effective in cognitive rehabilitation for patients affected by stroke [34,36,56]. However, our pooled results showed no significant benefits of VR in improving global cognitive function for patients with stroke. Our result was comparable to recently published meta-analysis reviews by Wiley et al [39] and Zhang et al [40], who reported that VR therapy is not superior to control interventions in improving global cognition in individuals with stroke. This result differs from another systematic review [25], which only included 4 studies published in 2018. The review found that there is a small-to-medium effect of VR therapy on cognitive outcomes for people after a stroke. A possible explanation for this conflicting finding is that most VR-based therapies do not always concentrate on the training of cognitive function. It remains to be determined whether more training sessions specifically focused on global cognitive function would affect outcomes.

Furthermore, the samples had relatively heterogeneous characteristics, such as stroke onset, the severity of the impairment, and lesion characteristics, which may affect cognition recovery in poststroke patients.

Cognitive function is a complex concept that includes various domain-specific cognition, such as attention, executive function, memory, and psychomotor speed [61]. In our review, we found improvements in executive function, memory, and visuospatial ability, indicating that VR can be considered an effective therapy for improving domain-specific cognition. These positive findings of VR therapy are in line with previous reviews [19,62,63], which report that VR represents a promising methodological approach, implementing specific cognitive and behavioural functions, such as executive function, attention, spatial cognition, memory, and language. A previous systematic review by Alexander et al [64] also supports these positive findings on domain-specific cognition. Wiley et al [39] employed a systematic review to examine the effectiveness of exercise-based VR therapy for poststroke, with 8 studies involving 196 participants, which did not find improvements in attention, memory, and language poststroke. The added value of VR in domain-specific cognition compared with most of the currently provided conventional therapies may be underlined by the following mechanisms. Flannery et al [65] stated that VR training activates brain metabolism, increases cerebral blood flow, and the release of neurotransmitters. Carrieri et al [66] confirmed that VR can foster the reactivation and improvement of various cortex functions and optimize the efficiency of the sensory cortex and is also effective in improving cognitive function.

Poor mental health after a stroke is common and complex. An individual's QoL is greatly reduced by re-experiencing the situation with depression, anxiety, agitation, and stress. Many studies and clinical trials have shown the potential of VR in relieving stress, depression, and anxiety in an imagined space, which makes it possible to provide efficient educational and psychological training without causing harm to patients with psychological problems [36,67]. There are several possible explanations for these improvements in mental health. Damsbo [68] explained that the use of VR allows individuals to learn emotion regulation strategies in the context of life-like virtual environments, thereby alleviating their negative emotions during long-term recovery. Choi et al [44] thought VR, through quasi-naturalistic and realistic stimuli in a multisensory fashion, can significantly enhance patients' awareness of the movement performed, as well as self-identification and self-recognition, and makes patients derive a high level of interest and enjoyment via the use of VR. However, the overall effects of VR on depression and the QoL were not encouraging in our meta-analysis. Some studies agree with the findings of our review [69,70]. The results of Lee [37] showed that some VR-based games are not suitable for patients following a stroke. Furthermore, patients with cognitive impairment may feel frustrated during the therapy, resulting in poor emotional experiences. Hence, the degree to which participants feel motivated and engaged during VR therapy can depend on the individual characteristics and the intervention content. The investigation of the mechanisms underlying how VR can influence mental functions is a critical point in stroke rehabilitation research. A likely cause of disagreement in findings on the QoL is that the QoL may worsen over time after stroke. However, the mean duration since stroke is inconsistent, even with no restrictions on duration among the trials included

in our review. It is, therefore, possible that the duration since stroke is an important factor that should be considered when selecting a VR therapy to increase the QoL for stroke patients.

Subgroup Analyses

Most subgroup analyses did not reveal significant differences between groups. Subgroup analyses consisted of a small number of comparisons, and the lack of relevant differences in most subgroup analyses might be caused by low statistical power. Only the subgroup analysis on global cognitive function showed that the effect of the usual-care control group was significant. There was also the issue with moderate-to-high heterogeneity in our meta-analysis, and grouping the included trials by measurement tools, duration of intervention, and delivery of control did not eliminate heterogeneity. Given the highly variable interventions of included trials, it is not surprising to find high heterogeneity in this case.

Strengths and Limitations

This meta-analysis followed the guidelines for performing rigorous systematic reviews [41]. Our review was only based on RCTs, which reinforces the evidence of our results. We also proposed a rigorous screening and search strategy to identify the most comprehensive literature in five major databases. Consequently, the results of our review are a widespread belief. However, there were still some potential limitations to our review. First, the number of trials included in our meta-analysis was limited. Hence, the power to detect small effect sizes was limited. Second, the included trials were extremely varied in terms of intervention contents and doses, and delivery of the control and measurement instruments, which made it rather complex to determine the optimal intensity of VR programs. Third, the quality, quantity, and sample size of included trials were far from ideal. Concealed treatment allocation was often not guaranteed, and blinding of outcome assessments was often not carried out. For example, only less than half of the included trials reported details of allocation concealment, leading to selection bias or confounding of the pooled results. Concerning detecting bias, almost half of the included trials failed to report information or had a high risk of blinding outcome assessment. Despite the rigorous nature of the included research designs (only included RCTs), our results must be interpreted with caution. Fourth, since we only considered effect post-intervention, the sustainability of VR-based therapy effects was not explored in our review. Indeed, details of long-term effects were unavailable in most trials included. Further research is needed to assess whether the effect observed in our review would persist over time and, if not, to evaluate at what duration and frequency the VR therapy should be repeated in order to sustain its effect. Finally, although a significant amount of work has been done in this area with promising results, the relevant characteristics of VR systems and the quantification of their impact on recovery are not yet clearly understood. As a result, we do not know how the different parameters of the proposed VR scenarios exactly affect recovery or whether they are effective at all. There is also a need to consider individual variability in order to optimize the impact of training.

Implications for Clinical Practice and Future Research

Notwithstanding potential limitations, the present findings offer some implications for researchers and health practitioners. In this high-tech era, clinicians may have more options and alternatives by providing an interactive and visually stimulating approach for patients with stroke, especially for those who cannot easily access traditional rehabilitation methods. The following issues need to be considered when applying VR-based therapies. First, VR should be adapted to the patient's needs and characteristics in performing activities, tasks, and tests. Second, safety problems need to be highlighted for older adults with reduced vision or other sensory problems. It is critical to consider a methodology type and an interaction technique that will result in the safe implementation of a VR therapy for patients with sensory and cognitive impairment. Finally, the price of VR equipment should be considered to meet the needs of target populations. In addition, there is a need for further research in this field to promote cost-effective care.

Conclusions

VR is a promising approach and can be used effectively in clinical neurorehabilitation. Although existing studies are limited, this review demonstrated statistically significant effects of VR-based therapies on executive function, memory, and visuospatial function in patients with stroke, but not on global cognitive function, attention, verbal fluency, depression, and the QoL. Larger, multicenter RCTs are warranted to confirm these positive effects. The completion of high - quality trials will ultimately advance the knowledge about optimal cognitive and psychological rehabilitation strategies for patients with stroke.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 22 KB - jmir_v23i11e31007_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 43 KB - jmir_v23i11e31007_app2.docx]

Multimedia Appendix 3
Forest plot for VR on depression and QoL. VR: virtual reality; QoL: quality of life.
[DOCX File , 211 KB - jmir_v23i11e31007_app3.docx]

Multimedia Appendix 4
The Egger funnel plot of publication bias.
[DOCX File , 20 KB - jmir_v23i11e31007_app4.docx]

References

1. Zhang Q, Zhang K, Li M, Gu J, Li X, Li M, et al. Validity and reliability of the Mandarin version of the Treatment Burden Questionnaire among stroke patients in Mainland China. Fam Pract 2021 Jul 28;38(4):537-542. [doi: 10.1093/fampra/cmab004] [Medline: 33615372]
2. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. The Lancet 2014 Jan 18;383(9913):245-255 [FREE Full text] [doi: 10.1016/s0140-6736(13)61953-4] [Medline: 24449944]
3. Patel MD, Tilling K, Lawrence E, Rudd AG, Wolfe CDA, McKevitt C. Relationships between long-term stroke disability, handicap and health-related quality of life. Age Ageing 2006 May;35(3):273-279. [doi: 10.1093/ageing/afj074] [Medline: 16638767]
4. Donovan NJ, Kendall DL, Heaton SC, Kwon S, Velozo CA, Duncan PW. Conceptualizing functional cognition in stroke. Neurorehabil Neural Repair 2008;22(2):122-135. [doi: 10.1177/1545968307306239] [Medline: 17761809]
5. Al-Qazzaz NK, Ali SH, Ahmad SA, Islam S, Mohamad K. Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment. Neuropsychiatr Dis Treat 2014;10:1677-1691 [FREE Full text] [doi: 10.2147/NDT.S67184] [Medline: 25228808]
6. Nys GMS, van Zandvoort MJE, de Kort PLM, van der Worp HB, Jansen BPW, Algra A, et al. The prognostic value of domain-specific cognitive abilities in acute first-ever stroke. Neurology 2005 Mar 08;64(5):821-827. [doi: 10.1212/01.WNL.0000152984.28420.5A] [Medline: 15753416]
7. Nys GMS, Van Zandvoort MJE, De Kort PLM, Jansen BPW, Van der Worp HB, Kappelle LJ, et al. Domain-specific cognitive recovery after first-ever stroke: a follow-up study of 111 cases. J Int Neuropsychol Soc 2005 Nov;11(7):795-806. [doi: 10.1017/s1355617705050952] [Medline: 16519259]
8. Qian J, McDonough DJ, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: a systematic review. Int J Environ Res Public Health 2020 Jun 10;17(11):4133 [FREE Full text] [doi: 10.3390/ijerph17114133] [Medline: 32531906]
9. Boger EJ, Demain S, Latter S. Self-management: a systematic review of outcome measures adopted in self-management interventions for stroke. Disabil Rehabil 2013 Aug;35(17):1415-1428 [FREE Full text] [doi: 10.3109/09638288.2012.737080] [Medline: 23167558]
10. Doré I, Caron J. Mental health: concepts, measures, determinants. Sante Ment Que 2017;42(1):125-145 [FREE Full text] [Medline: 28792565]
11. Ayerbe L, Ayis S, Rudd AG, Heuschmann PU, Wolfe CDA. Natural history, predictors, and associations of depression 5 years after stroke: the South London Stroke Register. Stroke 2011 Jul;42(7):1907-1911. [doi: 10.1161/STROKEAHA.110.605808] [Medline: 21566241]
12. Almhdawi KA, Alazrai A, Kanaan S, Shyyab AA, Oteir AO, Mansour ZM, et al. Post-stroke depression, anxiety, and stress symptoms and their associated factors: a cross-sectional study. Neuropsychol Rehabil 2021 Aug 18;31(7):1091-1104. [doi: 10.1080/09602011.2020.1760893] [Medline: 32419606]
13. Towfighi A, Ovbiagele B, El Husseini N, Hackett ML, Jorge RE, Kissela BM, American Heart Association Stroke Council, Council on Cardiovascular Stroke Nursing, Council on Quality of Care Outcomes Research. Poststroke depression: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2017 Feb;48(2):e30-e43. [doi: 10.1161/STR.0000000000000113] [Medline: 27932603]
14. Baumeister H, Härter M. Prevalence of mental disorders based on general population surveys. Soc Psychiatry Psychiatr Epidemiol 2007 Jul;42(7):537-546. [doi: 10.1007/s00127-007-0204-1] [Medline: 17516013]
15. Gillen R, Tennen H, McKee TE, Gernert-Dott P, Affleck G. Depressive symptoms and history of depression predict rehabilitation efficiency in stroke patients. Arch Phys Med Rehabil 2001 Dec;82(12):1645-1649. [doi: 10.1053/apmr.2001.26249] [Medline: 11733876]
16. Schöttke H, Gerke L, Düsing R, Möllmann A. Post-stroke depression and functional impairments: a 3-year prospective study. Compr Psychiatry 2020 May;99:152171 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152171] [Medline: 32179262]
17. Dellazizzo L, Potvin S, Luigi M, Dumais A. Evidence on virtual reality-based therapies for psychiatric disorders: meta-review of meta-analyses. J Med Internet Res 2020 Aug 19;22(8):e20889 [FREE Full text] [doi: 10.2196/20889] [Medline: 32812889]

18. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):e17331 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
19. Riva G, Mancuso V, Cavedoni S, Stramba-Badiale C. Virtual reality in neurorehabilitation: a review of its effects on multiple cognitive domains. Expert Rev Med Devices 2020 Oct;17(10):1035-1061. [doi: 10.1080/17434440.2020.1825939] [Medline: 32962433]
20. Threapleton K, Drummond A, Standen P. Virtual rehabilitation: what are the practical barriers for home-based research? Digit Health 2016 Apr;2:2055207616641302 [FREE Full text] [doi: 10.1177/2055207616641302] [Medline: 29942551]
21. Jung T, Moorhouse N, Shi X, Amin MF. A virtual reality-supported intervention for pulmonary rehabilitation of patients with chronic obstructive pulmonary disease: mixed methods study. J Med Internet Res 2020 Jul 07;22(7):e14178 [FREE Full text] [doi: 10.2196/14178] [Medline: 32673224]
22. Smith V, Warty RR, Sursas JA, Payne O, Nair A, Krishnan S, et al. The effectiveness of virtual reality in managing acute pain and anxiety for medical inpatients: systematic review. J Med Internet Res 2020 Nov 02;22(11):e17980 [FREE Full text] [doi: 10.2196/17980] [Medline: 33136055]
23. Kim H, Lee S, Jung D, Hur J, Lee H, Lee S, et al. Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: longitudinal questionnaire study. J Med Internet Res 2020 Oct 06;22(10):e23024 [FREE Full text] [doi: 10.2196/23024] [Medline: 33021481]
24. Kim W, Cho S, Ku J, Kim Y, Lee K, Hwang H, et al. Clinical application of virtual reality for upper limb motor rehabilitation in stroke: review of technologies and clinical evidence. J Clin Med 2020 Oct 21;9(10):3369 [FREE Full text] [doi: 10.3390/jcm9103369] [Medline: 33096678]
25. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. J Neuroeng Rehabil 2018 Mar 27;15(1):29 [FREE Full text] [doi: 10.1186/s12984-018-0370-2] [Medline: 29587853]
26. Jin R, Pilozzi A, Huang X. Current cognition tests, potential virtual reality applications, and serious games in cognitive assessment and non-pharmacological therapy for neurocognitive disorders. J Clin Med 2020 Oct 13;9(10):3287 [FREE Full text] [doi: 10.3390/jcm9103287] [Medline: 33066242]
27. Imbimbo I, Coraci D, Santilli C, Loreti C, Piccinini G, Ricciardi D, et al. Parkinson's disease and virtual reality rehabilitation: cognitive reserve influences the walking and balance outcome. Neurol Sci 2021 Mar 04. [doi: 10.1007/s10072-021-05123-3] [Medline: 33661481]
28. Shen J, Xiang H, Luna J, Grishchenko A, Patterson J, Strouse RV, et al. Virtual reality-based executive function rehabilitation system for children with traumatic brain injury: design and usability study. JMIR Serious Games 2020 Aug 25;8(3):e16947 [FREE Full text] [doi: 10.2196/16947] [Medline: 32447275]
29. Teo W, Muthalib M, Yamin S, Hendy AM, Bramstedt K, Kotsopoulos E, et al. Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation? A narrative review of the literature. Front Hum Neurosci 2016 Jun 24;10:284 [FREE Full text] [doi: 10.3389/fnhum.2016.00284] [Medline: 27445739]
30. Edmans JA, Gladman JRF, Cobb S, Sunderland A, Pridmore T, Hilton D, et al. Validity of a virtual environment for stroke rehabilitation. Stroke 2006 Nov;37(11):2770-2775. [doi: 10.1161/01.STR.0000245133.50935.65] [Medline: 17008609]
31. Levin MF. Can virtual reality offer enriched environments for rehabilitation? Expert Rev Neurother 2011 Feb;11(2):153-155. [doi: 10.1586/ern.10.201] [Medline: 21306202]
32. Taylor MJD, McCormick D, Shawis T, Impson R, Griffin M. Activity-promoting gaming systems in exercise and rehabilitation. J Rehabil Res Dev 2011;48(10):1171-1186 [FREE Full text] [doi: 10.1682/jrrd.2010.09.0171] [Medline: 22234662]
33. Gamito P, Oliveira J, Coelho C, Morais D, Lopes P, Pacheco J, et al. Cognitive training on stroke patients via virtual reality-based serious games. Disabil Rehabil 2017 Feb 05;39(4):385-388. [doi: 10.3109/09638288.2014.934925] [Medline: 25739412]
34. Kim BR, Chun MH, Kim LS, Park JY. Effect of virtual reality on cognition in stroke patients. Ann Rehabil Med 2011 Aug;35(4):450-459 [FREE Full text] [doi: 10.5535/arm.2011.35.4.450] [Medline: 22506159]
35. Cho K, Yu J, Jung J. Effects of virtual reality-based rehabilitation on upper extremity function and visual perception in stroke patients: a randomized control trial. J Phys Ther Sci 2012;24(11):1205-1208. [doi: 10.1589/jpts.24.1205] [Medline: 33050396]
36. Lin R, Chiang S, Heitkemper MM, Weng S, Lin C, Yang F, et al. Effectiveness of early rehabilitation combined with virtual reality training on muscle strength, mood state, and functional status in patients with acute stroke: a randomized controlled trial. Worldviews Evid Based Nurs 2020 Apr;17(2):158-167. [doi: 10.1111/wvn.12429] [Medline: 32212254]
37. Lee H, Huang C, Ho S, Sung W. The effect of a virtual reality game intervention on balance for patients with stroke: a randomized controlled trial. Games Health J 2017 Oct;6(5):303-311. [doi: 10.1089/g4h.2016.0109] [Medline: 28771379]
38. Lee CH, Kim YS, Jung JH. Effectiveness of virtual reality based cognitive rehabilitation on cognitive function, motivation and depression in stroke patients. Medico Legal Update 2020;20(1):1880-1886. [doi: 10.37506/mlu.v20i1.652]

39. Wiley E, Khattab S, Tang A. Examining the effect of virtual reality therapy on cognition post-stroke: a systematic review and meta-analysis. Disabil Rehabil Assist Technol 2020 May 02:1-11. [doi: 10.1080/17483107.2020.1755376] [Medline: 32363955]
40. Zhang B, Li D, Liu Y, Wang J, Xiao Q. Virtual reality for limb motor function, balance, gait, cognition and daily function of stroke patients: a systematic review and meta-analysis. J Adv Nurs 2021 Aug 06;77(8):3255-3273. [doi: 10.1111/jan.14800] [Medline: 33675076]
41. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. London: Cochrane; 2021.
42. Cho D, Lee S. Effects of virtual reality immersive training with computerized cognitive training on cognitive function and activities of daily living performance in patients with acute stage stroke: a preliminary randomized controlled trial. Medicine (Baltimore) 2019 Mar;98(11):e14752 [FREE Full text] [doi: 10.1097/MD.0000000000014752] [Medline: 30882644]
43. Choi D, Choi W, Lee S. Influence of Nintendo Wii Fit balance game on visual perception, postural balance, and walking in stroke survivors: a pilot randomized clinical trial. Games Health J 2018 Dec;7(6):377-384. [doi: 10.1089/g4h.2017.0126]
44. Choi H, Shin W, Bang D. Mirror therapy using gesture recognition for upper limb function, neck discomfort, and quality of life after chronic stroke: a single-blind randomized controlled trial. Med Sci Monit 2019 May 03;25:3271-3278 [FREE Full text] [doi: 10.12659/MSM.914095] [Medline: 31050660]
45. Choi JH, Han EY, Kim BR, Kim SM, Im SH, Lee SY, et al. Effectiveness of commercial gaming-based virtual reality movement therapy on functional recovery of upper extremity in subacute stroke patients. Ann Rehabil Med 2014 Aug;38(4):485-493 [FREE Full text] [doi: 10.5535/arm.2014.38.4.485] [Medline: 25229027]
46. Shin J, Bog Park S, Ho Jang S. Effects of game-based virtual reality on health-related quality of life in chronic stroke patients: a randomized, controlled study. Comput Biol Med 2015 Aug;63:92-98. [doi: 10.1016/j.compbiomed.2015.03.011] [Medline: 26046499]
47. Shin J, Kim M, Lee J, Jeon Y, Kim S, Lee S, et al. Effects of virtual reality-based rehabilitation on distal upper extremity function and health-related quality of life: a single-blinded, randomized controlled trial. J Neuroeng Rehabil 2016 Feb 24;13(1):17 [FREE Full text] [doi: 10.1186/s12984-016-0125-x] [Medline: 26911438]
48. Kim DH, Kim K, Lee S. The effects of effects of virtual reality training with upper limb sensory exercise stimulation on the AROM of upper limb joints, function, and concentration in chronic stroke patients. Phys Med Rehabil Kurortmed 2019 Jul 16;30(02):86-94. [doi: 10.1055/a-0917-4604]
49. Oh Y, Kim G, Han K, Won YH, Park S, Seo J, et al. Efficacy of virtual reality combined with real instrument training for patients with stroke: a randomized controlled trial. Arch Phys Med Rehabil 2019 Aug;100(8):1400-1408 [FREE Full text] [doi: 10.1016/j.apmr.2019.03.013] [Medline: 31002812]
50. Park M, Ko M, Oh S, Lee J, Ham Y, Yi H, et al. Effects of virtual reality-based planar motion exercises on upper extremity function, range of motion, and health-related quality of life: a multicenter, single-blinded, randomized, controlled pilot study. J Neuroeng Rehabil 2019 Oct 24;16(1):122 [FREE Full text] [doi: 10.1186/s12984-019-0595-8] [Medline: 31651335]
51. Faria AL, Andrade A, Soares L, I Badia SB. Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients. J Neuroeng Rehabil 2016 Nov 02;13(1):96 [FREE Full text] [doi: 10.1186/s12984-016-0204-z] [Medline: 27806718]
52. Faria AL, Pinho MS, Bermúdez I Badia S. A comparison of two personalization and adaptive cognitive rehabilitation approaches: a randomized controlled trial with chronic stroke patients. J Neuroeng Rehabil 2020 Jun 16;17(1):78 [FREE Full text] [doi: 10.1186/s12984-020-00691-5] [Medline: 32546251]
53. Faria AL, Cameirão MS, Couras JF, Aguiar JRO, Costa GM, Bermúdez I Badia S. Combined cognitive-motor rehabilitation in virtual reality improves motor outcomes in chronic stroke: a pilot study. Front Psychol 2018;9:854 [FREE Full text] [doi: 10.3389/fpsyg.2018.00854] [Medline: 29899719]
54. Unibaso-Markaida I, Iraurgi I, Ortiz-Marqués N, Amayra I, Martínez-Rodríguez S. Effect of the Wii Sports Resort on the improvement in attention, processing speed and working memory in moderate stroke. J Neuroeng Rehabil 2019 Feb 28;16(1):32 [FREE Full text] [doi: 10.1186/s12984-019-0500-5] [Medline: 30819204]
55. Ballester BR, Nirme J, Camacho I, Duarte E, Rodríguez S, Cuxart A, et al. Domiciliary VR-based therapy for functional recovery and cortical reorganization: randomized controlled trial in participants at the chronic stage post stroke. JMIR Serious Games 2017 Aug 07;5(3):e15 [FREE Full text] [doi: 10.2196/games.6773] [Medline: 28784593]
56. Rogers JM, Duckworth J, Middleton S, Steenbergen B, Wilson PH. Elements virtual rehabilitation improves motor, cognitive, and functional outcomes in adult stroke: evidence from a randomized controlled pilot study. J Neuroeng Rehabil 2019 May 15;16(1):56 [FREE Full text] [doi: 10.1186/s12984-019-0531-y] [Medline: 31092252]
57. Johnson L, Bird M, Muthalib M, Teo W. An innovative STRoke Interactive Virtual thErapy (STRIVE) online platform for community-dwelling stroke survivors: a randomized controlled trial. Arch Phys Med Rehabil 2020 Jul;101(7):1131-1137. [doi: 10.1016/j.apmr.2020.03.011] [Medline: 32283048]
58. Baltaduonienė D, Kubilius R, Berškienė K, Vitkus L, Petruševičienė D. Change of cognitive functions after stroke with rehabilitation systems. Transl Neurosci 2019;10:118-124 [FREE Full text] [doi: 10.1515/tnsci-2019-0020] [Medline: 31149357]

59. da Silva Ribeiro NM, Ferraz DD, Pedreira É, Pinheiro Í, da Silva Pinto AC, Neto MG, et al. Virtual rehabilitation via Nintendo Wii® and conventional physical therapy effectively treat post-stroke hemiparetic patients. Top Stroke Rehabil 2015 Aug;22(4):299-305. [doi: 10.1179/1074935714Z.0000000017] [Medline: 26258455]
60. Şimşek TT, Çekok K. The effects of Nintendo Wii(TM)-based balance and upper extremity training on activities of daily living and quality of life in patients with sub-acute stroke: a randomized controlled study. Int J Neurosci 2016 Dec;126(12):1061-1070. [doi: 10.3109/00207454.2015.1115993] [Medline: 26626539]
61. Zhang Q, Hu JN, Wei LJ. Effects of traditional Chinese exercise on cognitive and psychological outcomes in older adults with mild cognitive impairment: a systematic review and meta-analysis. Medicene 2019 Feb:e14581. [doi: 10.1097/md.0000000000014581] [Medline: 30762810]
62. Maggio MG, Latella D, Maresca G, Sciarrone F, Manuli A, Naro A, et al. Virtual reality and cognitive rehabilitation in people with stroke: an overview. J Neurosci Nurs 2019 Apr;51(2):101-105. [doi: 10.1097/jnn.0000000000000423] [Medline: 30649091]
63. Maggio MG, Maresca G, De Luca R, Stagnitti MC, Porcari B, Ferrera MC, et al. The growing use of virtual reality in cognitive rehabilitation: fact, fake or vision? A scoping review. J Natl Med Assoc 2019 Aug;111(4):457-463. [doi: 10.1016/j.jnma.2019.01.003] [Medline: 30739728]
64. Moreno A, Wall KJ, Thangavelu K, Craven L, Ward E, Dissanayaka NN. A systematic review of the use of virtual reality and its effects on cognition in individuals with neurocognitive disorders. Alzheimers Dement (N Y) 2019 Jan;5(1):834-850 [FREE Full text] [doi: 10.1016/j.trci.2019.09.016] [Medline: 31799368]
65. Flannery RB. Treating learned helplessness in the elderly dementia patient: preliminary inquiry. Am J Alzheimers Dis Other Demen 2002;17(6):345-349 [FREE Full text] [doi: 10.1177/153331750201700605] [Medline: 12501481]
66. Carrieri M, Petracca A, Lancia S, Basso Moro S, Brigadoi S, Spezialetti M, et al. Prefrontal cortex activation upon a demanding virtual hand-controlled task: a new frontier for neuroergonomics. Front Hum Neurosci 2016 Jul 19;10(7):53 [FREE Full text] [doi: 10.3389/fnhum.2016.00053] [Medline: 26909033]
67. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019 Jul;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
68. Damsbo AG, Kraglund KL, Buttenschøn HN, Johnsen SP, Andersen G, Mortensen JK. Predictors for wellbeing and characteristics of mental health after stroke. J Affect Disord 2020 Mar 01;264:358-364 [FREE Full text] [doi: 10.1016/j.jad.2019.12.032] [Medline: 32056772]
69. Lee SH, Lee J, Kim M, Jeon Y, Kim S, Shin J. Virtual reality rehabilitation with functional electrical stimulation improves upper extremity function in patients with chronic stroke: a pilot randomized controlled study. Arch Phys Med Rehabil 2018 Aug;99(8):1447-1453.e1. [doi: 10.1016/j.apmr.2018.01.030] [Medline: 29505744]
70. Choi Y, Ku J, Lim H, Kim YH, Paik N. Mobile game-based virtual reality rehabilitation program for upper limb dysfunction after ischemic stroke. Restor Neurol Neurosci 2016 May 02;34(3):455-463. [doi: 10.3233/RNN-150626] [Medline: 27163250]

Abbreviations

CI: confidence interval
LOTCA: Loewenstein Occupational Therapy Cognitive Assessment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference
VR: virtual reality
WMD: weighted mean difference

Edited by R Kukafka; submitted 06.06.21; peer-reviewed by JW Hur, HS Kim, J Wang; comments to author 11.08.21; revised version received 14.08.21; accepted 12.09.21; published 17.11.21.

Please cite as:
Zhang Q, Fu Y, Lu Y, Zhang Y, Huang Q, Yang Y, Zhang K, Li M
Impact of Virtual Reality-Based Therapies on Cognition and Mental Health of Stroke Patients: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(11):e31007
URL: https://www.jmir.org/2021/11/e31007
doi:10.2196/31007
PMID:34787571

Review

eHealth Literacy Instruments: Systematic Review of Measurement Properties

Jiyeon Lee[1,2], PhD; Eun-Hyun Lee[3], PhD; Duckhee Chae[4], PhD

[1]College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, Republic of Korea

[3]Graduate School of Public Health, Ajou University, Suwon, Republic of Korea

[4]College of Nursing, Chonnam National University, Gwangju, Republic of Korea

Corresponding Author:
Eun-Hyun Lee, PhD
Graduate School of Public Health
Ajou University
164, Worldcup-ro, Yeongtong-gu
Suwon, 16499
Republic of Korea
Phone: 82 31 219 5296
Fax: 82 31 219 5025
Email: ehlee@ajou.ac.kr

Abstract

Background: The internet is now a major source of health information. With the growth of internet users, eHealth literacy has emerged as a new concept for digital health care. Therefore, health professionals need to consider the eHealth literacy of consumers when providing care utilizing digital health technologies.

Objective: This study aimed to identify currently available eHealth literacy instruments and evaluate their measurement properties to provide robust evidence to researchers and clinicians who are selecting an eHealth literacy instrument.

Methods: We conducted a systematic review and meta-analysis of self-reported eHealth literacy instruments by applying the updated COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) methodology.

Results: This study included 7 instruments from 41 articles describing 57 psychometric studies, as identified in 4 databases (PubMed, CINAHL, Embase, and PsycInfo). No eHealth literacy instrument provided evidence for all measurement properties. The eHealth literacy scale (eHEALS) was originally developed with a single-factor structure under the definition of eHealth literacy before the rise of social media and the mobile web. That instrument was evaluated in 18 different languages and 26 countries, involving diverse populations. However, various other factor structures were exhibited: 7 types of two-factor structures, 3 types of three-factor structures, and 1 bifactor structure. The transactional eHealth literacy instrument (TeHLI) was developed to reflect the broader concept of eHealth literacy and was demonstrated to have a sufficient low-quality and very low-quality evidence for content validity (relevance, comprehensiveness, and comprehensibility) and sufficient high-quality evidence for structural validity and internal consistency; however, that instrument has rarely been evaluated.

Conclusions: The eHealth literacy scale was the most frequently investigated instrument. However, it is strongly recommended that the instrument's content be updated to reflect recent advancements in digital health technologies. In addition, the transactional eHealth literacy instrument needs improvements in content validity and further psychometric studies to increase the credibility of its synthesized evidence.

(J Med Internet Res 2021;23(11):e30644) doi:10.2196/30644

KEYWORDS

eHealth literacy; systematic review; meta-analysis; psychometrics; reliability; validity; scale; instrument

Introduction

Health literacy is an important determinant for achieving positive health outcomes [1-3]. It refers to the ability to "assess, understand, appraise and apply health information to make judgments and make decisions in everyday life concerning health care, disease prevention and health promotion (p. 3)" [4]. The primary sources for obtaining health information have previously been traditional media (eg, books, brochures, newspapers, and television) and the attending health professionals [5].

The internet is now a major source of health information [6]. There were 5.09 billion internet users worldwide in 2021, representing 64.7% of the global population [7]. In Europe, between 70% and 90% of internet users access health information [8], while about 72% of internet users in the United States search for health information on the internet [9]. Obtaining health information from the internet requires the skills to utilize digital technologies to search and acquire information and basic health literacy abilities such as reading, understanding, and appraising health information. This perspective resulted in the emergence of eHealth literacy in 2006. An early definition proposed for eHealth literacy was "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" (p.2) [10].

The rapidly increasing use of digital devices (eg, computers, tablets, and smartphones) and the internet means that health professionals are transiting the method of health information delivery beyond a traditional face-to-face mode into a web-based model, largely due to its advantages of not being restricted to time and space [11]. To ensure the effective web-based delivery of health information, health professionals need to consider the eHealth literacy of consumers. Due to the COVID-19 pandemic requiring quarantining and social isolation, face-to-face visiting of patients with chronic diseases became difficult; therefore, the use of remote care using digital health technologies was recommended as an alternative strategy for delivering health care and informational support [12]. As a result, assessments of eHealth literacy have accelerated as health professionals have attempted to adapt digital health services to patients.

The emergence of eHealth literacy has resulted in the development of self-reporting instruments to measure it. According to the United States Department of Health and Human Services [13], a newly developed or modified self-reporting instrument must satisfy certain measurement properties before applying it in practice or research. Using such an instrument without evidence regarding its measurement properties may misinform practitioners on the measuring concept and threaten the credibility of research results [14]. A systematic review of the measurement properties of eHealth literacy instruments could identify all existing instruments and provide psychometric information to determine which is the best.

One previous narrative review of eHealth literacy instruments [15] simply summarized instruments rather than performing quality assessments or data syntheses. The COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) is the most popular methodology for systematically reviewing measurement properties of self-reported instruments [16-18]. To the best of our knowledge, such a systematic review of the measurement properties of eHealth literacy instruments has not been conducted previously. Therefore, this study aimed (1) to identify the currently available instruments for measuring eHealth literacy and (2) to evaluate their measurement properties to provide robust evidence for researchers and clinicians to use when selecting instruments.

Methods

Design and Searching Strategy

A systematic review of self-reported instruments was conducted according to the updated COSMIN methodology. The PubMed, CINAHL, Embase, and PsycInfo databases were searched from their dates of inception up to March 3, 2021. A search strategy based on the COSMIN involved constructing search filters for the key elements of the construct of interest: population(s), type of instruments (eg, scale or questionnaire), and measurement properties (including inclusion and exclusion filters), and then combining them using AND and NOT Boolean operators. The search filter used for the construct of interest (ie, eHealth literacy) in this study is presented in Multimedia Appendix 1. The search filter for population(s) was not applied because our study aimed to review all self-reported eHealth literacy instruments without considering specific populations. Regarding the type of instruments and the measurement properties, a modified filter developed by the Patient-Reported Outcomes Measurement Group at the University of Oxford and a validated highly sensitive search filter developed using the COSMIN were used [19].

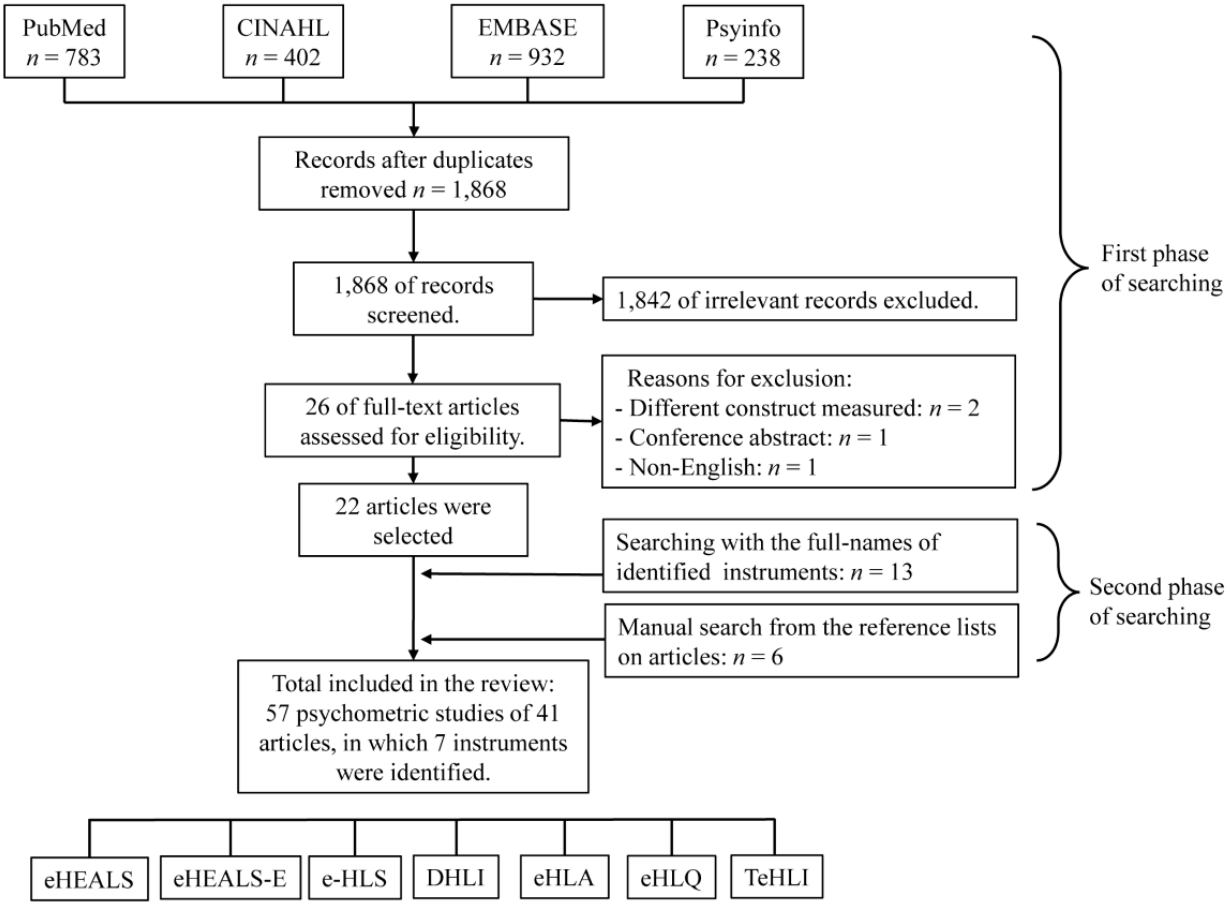
Eligibility Criteria

We included studies involving the development of an eHealth literacy instrument or evaluations of its measurement properties published as full-text original articles in peer-reviewed journals written in English. If a study had utilized an eHealth literacy instrument as an outcome measure and determined its measurement properties, such as Cronbach's α, but not with the main purpose of evaluating measurement properties of an eHealth literacy instrument, then the article was not included. Literature providing limited information such as conference abstracts, review protocols, or a note were also excluded.

Selection of Articles

Figure 1 presents a flow diagram of PRISMA (preferred reporting items for systematic reviews and meta-analyses) [20]. Duplicated records were removed using EndNote X8.2 (Thomas Reuters). Two reviewers (JL and DC) independently selected articles based on their abstracts and full texts. Differences were discussed, and a consensus was reached by consulting with the third reviewer (E-HL). After identifying an initial list of articles and included instruments in the first phase of searching, database searching utilizing the full names of the identified instruments and the measurement-property filter was conducted in the second phase, which also included manual searching based on the reference lists of the selected articles.

Figure 1. PRISMA flow diagram. DHLI: digital health literacy instrument; eHEALS: eHealth literacy scale; eHEALS-E: eHealth literacy scale–extended; eHLA: eHealth literacy assessment toolkit; e-HLS: electronic health literacy scale; eHLQ: eHealth literacy questionnaire; PRISMA: preferred reporting items for systematic reviews and meta-analyses; TeHLI: transactional eHealth literacy instrument.



Data Extraction

Data were extracted from each article to understand the characteristics of the analyzed instrument (ie, target population, number of subscales and items, response options, mode of administration, and language used for the instrument), the study samples (ie, sample size, age, gender) used to assess the identified instruments, theoretical/conceptual frameworks and specified definitions used for the development of instruments, and the results of measurement properties and floor and ceiling effects of the eHealth literacy instruments.

Evaluating the Measurement Properties of the Instruments

The measurement properties of the instruments were evaluated in 3 steps. First, the methodological quality of the included studies was evaluated using the COSMIN Risk of Bias checklist [16,18]. Each measurement property in each study was evaluated using items in the checklist and rated as very good, adequate, doubtful, or inadequate. The lowest rating of any standard in the box was taken as the methodological quality. Regarding the evaluation of each measurement property, content validity was the first parameter to be evaluated. Content validity (relevance, comprehensiveness, and comprehensibility) was considered the most important measurement property because an instrument needs to reflect the construct being measured adequately. Next, the internal structure of an instrument (structural validity, internal consistency, and cross-cultural validity or measurement invariance) was evaluated. The structural validity of the instrument such as a one-factor or two-factor structure guided the evaluation of internal consistency; for example, when a one-factor structure was supported, then Cronbach's α for all items needed to be evaluated, whereas if a two-factor structure was supported, we needed to evaluate the Cronbach's α of two subscales. Subsequently, remaining measurement properties such as reliability, measurement error, hypotheses testing for construct validity (convergent validity and discriminant or known-groups validity), and responsiveness were evaluated. The methodological quality of criterion validity was not evaluated since there is no gold standard for eHealth literacy measures.

Second, the results of each study for measurement properties were rated according to the updated quality criteria for good measurement properties as sufficient (+), insufficient (–), or indeterminate (?) [18,21]. The quality criteria use only Cronbach's α (≥.70) as the rating indicator of internal consistency. Therefore, the following internal consistency-related criteria were added: (1) sufficient (+) for an omega or person/item reliability of ≥.70 for each unidimensional scale or subscale, insufficient (–) for an omega or person/item reliability of <.70, and indeterminate (?) if the

values were not reported; and (2) sufficient (+) for a person/item separation index of ≥1.50 for each unidimensional scale or subscale, insufficient (–) for a person/item separation index of <1.50, and indeterminate (?) if the values were not reported [22]. Additional criteria suggested by Lee et al [23] were applied to evaluate the structural validity obtained in exploratory factor analysis (eg, factor explanation of at least 50% of the variance). The criterion for hypotheses testing (convergent validity) was set as $r \geq .30$ with other comparators such as health literacy and internet-related and health-related variables (eg, internet use and adherence to a regimen).

Finally, all of the results for each instrument's measurement properties were qualitatively summarized or quantitatively pooled through meta-analysis using statistical package meta in R software (version 4.0.3; R Core Team). The summarized results related to content validity were rated as sufficient (+), insufficient (–), or inconsistent (±) according to the 10 criteria for good content validity [18]. The summarized or pooled results for other properties were rated as sufficient (+), insufficient (–), inconsistent (±), or indeterminate (?) according to the quality criteria for good measurement properties [17]. Next, the quality of evidence for the overall ratings was graded as high, moderate, low, or very low using the modified GRADE (grading of recommendations assessment, development, and evaluation) approach considering the risk of bias, inconsistency, imprecision, and indirectness [17]. The quality of evidence for structural validity was a prerequisite for analyzing the internal consistency, and so it was taken as a starting point for determining the quality of evidence for internal consistency. The above processes were conducted by all 3 reviewers, with a consensus reached through discussion.

Results

Identified eHealth Literacy Instruments

The database search identified 2355 records (783, 402, 932, and 238 in PubMed, CINAHL, Embase, and PsycInfo, respectively), and 1868 records were screened after removing duplicates (Figure 1). In the first phase of searching, 22 articles were selected based on their titles and abstracts. Thirteen articles were identified in the second phase of database searching using the names of the identified instruments and the measurement-property filter, with 6 articles identified through manual searching of the reference lists of the selected articles. Therefore, the total number of included articles was 41. According to the COSMIN, each structure of an instrument is considered a separate study [16]. Some of the identified articles included multiple different factor structures; therefore, 57 studies in 41 articles were finally included in the present systematic review. The following seven instruments were identified: eHealth literacy scale (eHEALS), eHealth literacy scale–extended (eHEALS-E), electronic health literacy scale (e-HLS), digital health literacy instrument (DHLI), eHealth literacy assessment toolkit (eHLA), eHealth literacy questionnaire (eHLQ), and transactional eHealth literacy instrument (TeHLI).

Characteristics of the Included Instruments and Studies

The characteristics of the included eHealth literacy instruments and studies are presented in Multimedia Appendix 2 [24-64]. The eHEALS, which consists of 8 items scored on a 5-point Likert scale, was originally developed in English [24], and its psychometrics have been studied in diverse languages: Amharic [50], Mandarin Chinese [26], Simplified Chinese [34,36,41], Dutch [25], English [27-30,47,53,54,57,58], German [39,44], Greek [48], Hebrew [45], Hungarian [37], Indonesian [43], Italian [31,33], Korean [32,42,55], Persian [38,46], Polish [35], Portuguese [52], Norwegian [49,56], Serbian [51], and Swedish [40]. The eHEALS has been used to evaluate diverse populations, including not only youths, adults, and older adults, but also healthy people, patients, caregivers, and health professionals in school, community, and clinic settings. The recall period for the eHEALS was specified as “right now,” whereas other instruments did not specify recall periods.

The eHEALS-E is the extended version of the eHEALS comprising 20 items developed for the users of online health communities [59]. The e-HLS, with 19 items, was developed in the United States for online administration to the general population [60]. The DHLI has 21 items scored on a 4-point Likert scale originally developed in Dutch and English and targeting the general population [61]. In addition, this instrument has 3 items that are not obligatory to answer when respondents do not have experience posting messages on social media (ie, they can leave the items blank). The DHLI was further assessed in the Korean language for older adults in welfare centers [42]. The eHLA is the longest instrument, comprising 42 items scored on a 4-point Likert scale and using multiple choices, and was developed in both Danish and English [62]. The eHLQ comprises 35 items scored on a 4-point Likert scale and was also developed in Danish and English [63]. The TeHLI was developed in the United States for patients with lung disease and is composed of 18 items for online administration [64].

Theoretical/Conceptual Framework, Definition, and Intended Use

The theoretical/conceptual frameworks, definitions used when developing the identified instruments, and intended use are summarized in Table 1. The eHEALS and TeHLI were developed based on the Lily model and self-efficacy theory [10,65] and the transactional model of eHealth literacy (TMeHL) [66], respectively, and their specified definitions of eHealth literacy have been clarified. Both the eHLA and eHLQ were developed based on the eHLF (eHealth literacy framework) [67].

Table 1. Theoretical/conceptual framework, specified definition, and intended use.

Instrument	Authors	Theoretical/conceptual framework	Specified definition for the development of the instrument	Intended use
eHEALS[a]	Norman & Skinner [24]	Six components of the Lily model: traditional, computer, information, health, media, and science literacies [10]. Social cognitive theory (self-efficacy theory) [65].	"…the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem (p. 2)" [10].	"…designed to provide a general estimate of consumer eHealth-related skills" (p. 2) [10].
eHEALS-E[b]	Petri et al [59]	—[c]	(Additional items deduced from the definition of the concept used for the eHEALS development were included.)	"…accessing, understanding, appraising, and applying health-related online information" (p. 3) [59].
e-HLS[d]	Seçkin et al [60]	eHealth literacy was grounded on the construct of health literacy, and the three domains of trust, action, and behavior were identified in the literature.	—	"…designed to assess the degree to which people possess the skills required to use eHealth information in an informed way" (p. 3) [60].
DHLI[e]	van der Vaart & Drossaert [61]	The construct of eHealth literacy was derived from formative research of the actual performance tests [68].	—	"…to assess both Health 1.0 and Health 2.0 skills, using self-reporting and performance-based items" (p. 9) [61].
eHLA[f]	Karnoe et al [62]	The constructs of eHealth literacy were from the Lily model as well as the eHLF describing the interaction domains and their relations with individual and system domains [10,67].	—	"…suitable for screening purposes…" (p. 2) [62].
eHLQ[g]	Kayser et al. [63]	Seven-dimension eHLF[h] [67].	—	"…to support researchers, developers, designers, and governments to develop, implement, and evaluate effective digital health interventions" (p. 7) [63].
TeHLI[i]	Paige et al [64]	TMeHL[j] [66].	"The ability to locate, understand, exchange, and evaluate health information from online environments in the presence of dynamic contextual factors, and to apply the knowledge gained across ecological levels for the purposes of maintaining or improving health (p. 9)." [66]	"… to measure perceived skills related to the capacity to understand, exchange, evaluate, and apply health information from online multimedia" (p. 738) [64].

[a]eHEALS: eHealth literacy scale.

[b]eHEALS-E: eHealth literacy scale-extended.

[c]Cells left blank if no information was available in the study.

[d]e-HLS: electronic health literacy scale.

[e]DHLI: digital health literacy instrument.

[f]eHLA: eHealth literacy assessment toolkit.

[g]eHLQ: eHealth literacy questionnaire.

[h]eHLF: eHealth literacy framework.

[i]TeHLI: transactional eHealth literacy instrument.

[j]TMeHL: transactional model of eHealth literacy.

Overall Rating and Quality of Evidence for the Content Validity of Each Instrument

Table 2 presents the overall rating and quality of evidence for content validity for each instrument. The eHEALS was rated as having sufficient moderate-quality evidence for comprehensibility, whereas there was inconsistent low-quality evidence for relevance and insufficient very low-quality evidence for comprehensiveness. The eHEALS-E was rated as having inconsistent moderate-quality evidence for relevance, sufficient very low-quality evidence for comprehensiveness, and inconsistent very low-quality evidence for comprehensibility. The e-HLS, DHLI, eHLA, eHLQ, and TeHLI received sufficient ratings for relevance, comprehensiveness, and comprehensibility with low-quality or very low-quality evidence.

Table 2. Overall rating and quality of evidence for the content validity of each instrument.[a]

Instrument	Relevance		Comprehensiveness		Comprehensibility	
	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
eHEALS[b]	±	Low	−	Very low	+	Moderate
eHEALS-E[c]	±	Moderate	+	Very low	±	Very low
e-HLS[d]	+	Low	+	Low	+	Low
DHLI[e]	+	Low	+	Very low	+	Very low
eHLA[f]	+	Low	+	Low	+	Low
eHLQ[g]	+	Low	+	Low	+	Low
TeHLI[h]	+	Low	+	Very low	+	Low

[a]Sufficient (+), insufficient (−), and inconsistent (±).

[b]eHEALS: eHealth literacy scale.

[c]eHEALS-E: eHealth literacy scale-extended.

[d]e-HLS: electronic health literacy scale.

[e]DHLI: digital health literacy instrument.

[f]eHLA: eHealth literacy assessment toolkit.

[g]eHLQ: eHealth literacy questionnaire.

[h]TeHLI: transactional eHealth literacy instrument.

Overall Ratings and Quality of Evidence for Other Measurement Properties of Each Instrument

The measurement error and responsiveness were not assessed for any of the instruments; therefore, the results for structural validity, internal consistency, cross-cultural/measurement invariance, reliability, and hypotheses testing (convergent validity and discriminant/known-groups validity) were summarized or pooled for each instrument. The summarized or pooled results for the measurement properties of each instrument are presented in Multimedia Appendix 3. The overall rating and quality of evidence for the properties are presented in Tables 3 and 4.

The single-factor structure of the eHEALS (ID, study identification numbers 1-29) [24-43] demonstrated insufficient moderate-quality evidence (62.1% of the results supported the single-factor structure). Internal consistency of the single-factor eHEALS was supported through a meta-analysis with a Cronbach's α of 0.91 (Figure 2), as well as a qualitative summary with an omega of 0.89-0.94, person reliability of 0.80-0.87, person separation index of 2.36, item reliability index of 0.89-0.93, and item separation index of 3.62-11.3, which were rated as sufficient and indicated that there existed multiple studies of very good quality. According to the COSMIN, the quality of evidence for internal consistency cannot be greater than the quality of evidence for structural validity. Therefore, the quality of evidence for internal consistency was downgraded to moderate to reflect the quality of evidence for structural validity. Measurement invariance for parameters such as gender and age were evaluated in 5 studies and rated as sufficient high-quality evidence. Reliability and hypothesis testing for convergent validity demonstrated that there was insufficient high-quality evidence. There was sufficient moderate-quality evidence for discriminant/known-groups validity.

The second-most-frequent structure of the eHEALS was a two-factor structure. However, the subscale structures were not identical. A two-factor structure as derived from 3 studies (IDs 30-32) [31,39,44] demonstrated insufficient high-quality evidence for structural validity and sufficient high-quality evidence for internal consistency. However, there was inconsistent moderate-quality evidence for convergent validity. The two-factor structure yielded from another 5 studies (IDs 35-39) [47-50] demonstrated insufficient high-quality evidence for structural validity, sufficient high-quality evidence for internal consistency, and sufficient very low-quality evidence for reliability and convergent validity. The three-factor structure of the eHEALS derived from 3 studies (IDs 43-45) [54-56] and a single study ID 47 [58] demonstrated sufficient high-quality evidence for internal consistency, cross-cultural validity, and known-groups validity. The eHEALS derived from 3 studies (IDs 43-45) [54-56] demonstrated insufficient low-quality evidence for reliability and insufficient high-quality evidence for convergent validity, whereas the eHEALS derived from a single study ID 47 [58] did not evaluate these properties; thus no evidence existed.

Table 3. Overall rating and quality of evidence for measurement properties of structural validity, internal consistency, and cross-cultural/measurement invariance.[a]

Study ID[b]	Instrument	# of factors	Structural validity		Internal consistency		Cross-cultural/ measurement invariance	
			Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
1-29 [24-43]	eHEALS[c]	1	–	Moderate	+	Moderate	+	High
30-32[31,39,44]	eHEALS	2[d]	–	High	+	High	N/A[e]	N/A
33 [45]	eHEALS	2[f]	+	Low	+	Low	N/A	N/A
34 [46]	eHEALS	2[g]	+	Moderate	N/A	N/A	N/A	N/A
35–39 [47-50]	eHEALS	2[h]	–	High	+	High	N/A	N/A
40 [51]	eHEALS	2[i]	+	Moderate	N/A	N/A	N/A	N/A
41 [52]	eHEALS	2[j]	+	Moderate	+	Moderate	N/A	N/A
42 [53]	eHEALS	2[k]	+	Moderate	+	Moderate	N/A	N/A
43-45 [54-56]	eHEALS	3[l]	+	High	+	High	+	High
46 [57]	eHEALS	3[m]	+	Low	N/A	N/A	N/A	N/A
47 [58]	eHEALS	3[n]	+	High	+	High	+	High
48 [39]	eHEALS	Bifactor[o]	?	Low	N/A	N/A	N/A	N/A
49 [59]	eHEALS-E[p]	6	+	High	+	High	N/A	N/A
50 [60]	e-HLS[q]	3	–	Low	N/A	N/A	N/A	N/A
51 [61]	DHLI[r]	7	+	Low	+	Low	N/A	N/A
52 [42]	DHLI	5	+	Low	N/A	N/A	N/A	N/A
53 [62]	eHLA[s]	7	?	Very low	–	Very low	N/A	N/A
54, 55 [63]	eHLQ[t]	7	–	High	+	High	?	Low
56, 57 [64]	TeHLI[u]	4	+	High	+	High	N/A	N/A

[a]The item numbers of the eHEALS are those assigned in the original article by Norman and Skinner [24].

[b]ID: study identification number (a study identification number was assigned to each of the 57 studies in the 41 articles because some articles covered multiple studies; see Multimedia Appendix 2).

[c]eHEALS: eHealth literacy scale.

[d]Information seeking (items 1, 2, 3, 4, 5, 8), information appraisal (items 6, 7) [31,39,44].

[e]No information was available in the study.

[f]Factor 1 (items 1, 2, 4), factor 2 (items 3, 5, 6, 7, 8) [45].

[g]Factor 1 (items 3, 4), factor 2 (items 1, 2, 5, 6, 7, 8) [46].

[h]Factor 1 (items 1, 2, 3, 4, 5), factor 2 (items 6, 7, 8) [47-50].

[i]Factor 1 (items 2, 6, 7, 8), factor 2 (items 1, 3, 4, 5) [51].

[j]Factor 1 (items 1, 2, 3, 4), factor 2 (items 5, 6, 7, 8) [52].

[k]Information acquisition (items 1, 3, 4), information application (items 2, 5, 6, 7, 8) [53].

[l]Awareness (items 3, 4), skills (items 1, 2, 5), evaluation (items 6, 7, 8) [54-56].

[m]Awareness (items 1, 2), skills (items 4, 5), evaluation (items 6, 7, 8) [57].

[n]Information awareness (items 3, 4), information seeking (items 1, 5), information engagement (items 2, 6, 7, 8) [58].

[o]General factor (items 1, 2, 3, 4, 5, 6, 7, 8), subfactor 1 (items 1, 2, 3, 4, 5, 8), subfactor 2 (items 6, 7) [39].

[p]eHEALS-E: eHealth literacy scale-extended.

[q]e-HLS: electronic health literacy scale.

[r]DHLI: digital health literacy instrument

[s]eHLA: eHealth literacy assessment toolkit.

[t]eHLQ: eHealth literacy questionnaire.

[u]TeHLI: transactional eHealth literacy instrument.

Table 4. Overall rating and quality of evidence for measurement properties of reliability, convergent validity, and discriminant/known-groups validity.[a]

Study ID[b]	Instrument	No. of factors	Reliability		Hypothesis testing: convergent validity		Hypothesis testing: discriminant/known-groups validity	
			Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
1-29 [24-43]	eHEALS[c]	1	–	High	–	High	+	Moderate
30-32[31,39,44]	eHEALS	2[d]	N/A[e]	N/A	±	Moderate	N/A	N/A
33 [45]	eHEALS	2[f]	N/A	N/A	+	Moderate	N/A	N/A
34 [46]	eHEALS	2[g]	–	Very low	?	Very low	N/A	N/A
35-39 [47-50]	eHEALS	2[h]	+	Very low	+	Very low	N/A	N/A
40 [51]	eHEALS	2[i]	N/A	N/A	N/A	N/A	N/A	N/A
41 [52]	eHEALS	2[j]	N/A	N/A	N/A	N/A	–	Low
42 [53]	eHEALS	2[k]	N/A	N/A	–	Moderate	N/A	N/A
43-45[54-56]	eHEALS	3[l]	–	Low	–	High	+	High
46 [57]	eHEALS	3[m]	N/A	N/A	N/A	N/A	N/A	N/A
47 [58]	eHEALS	3[n]	N/A	N/A	N/A	N/A	+	High
48 [39]	eHEALS	Bifactor[o]	N/A	N/A	N/A	N/A	N/A	N/A
49 [59]	eHEALS-E[p]	6	N/A	N/A	N/A	N/A	–	Low
50 [60]	e-HLS[q]	3	N/A	N/A	–	Very low	N/A	N/A
51 [61]	DHLI[r]	7	+	Low	–	High	N/A	N/A
52 [42]	DHLI	5	+	Low	–	Low	N/A	N/A
53 [62]	eHLA[s]	7	N/A	N/A	N/A	N/A	N/A	N/A
54, 55 [63]	eHLQ[t]	7	N/A	N/A	N/A	N/A	N/A	N/A
56, 57 [64]	TeHLI[u]	4	N/A	N/A	±	Low	N/A	N/A

[a]The item numbers of the eHEALS are those assigned in the original article by Norman and Skinner [24].

[b]ID: study identification number (a study identification number was assigned to each of the 57 studies in the 41 articles because some articles covered multiple studies; see Multimedia Appendix 2).

[c]eHEALS: eHealth literacy scale.

[d]Information seeking (items 1, 2, 3, 4, 5, 8), information appraisal (items 6, 7) [31,39,44].

[e]No information was available in the study.

[f]Factor 1 (items 1, 2, 4), factor 2 (items 3, 5, 6, 7, 8) [45].

[g]Factor 1 (items 3, 4), factor 2 (items 1, 2, 5, 6, 7, 8) [46].

[h]Factor 1 (items 1, 2, 3, 4, 5), factor 2 (items 6, 7, 8) [47-50].

[i]Factor 1 (items 2, 6, 7, 8), factor 2 (items 1, 3, 4, 5) [51].

[j] Factor 1 (items 1, 2, 3, 4), factor 2 (items 5, 6, 7, 8) [52].

[k]Information acquisition (items 1, 3, 4), information application (items 2, 5, 6, 7, 8) [53].

[l]Awareness (items 3, 4), skills (items 1, 2, 5), evaluation (items 6, 7, 8) [54-56].

[m]Awareness (items 1, 2), skills (items 4,5), evaluation (items 6, 7, 8) [57].

[n]Information awareness (items 3, 4), information seeking (items 1, 5), information engagement (items 2, 6, 7, 8) [58].

[o]General factor (items 1, 2, 3, 4, 5, 6, 7, 8), subfactor 1 (items 1, 2, 3, 4, 5, 8), subfactor 2 (items 6, 7) [39].

[p]eHEALS-E: eHealth literacy scale-extended.

[q]e-HLS: electronic health literacy scale.

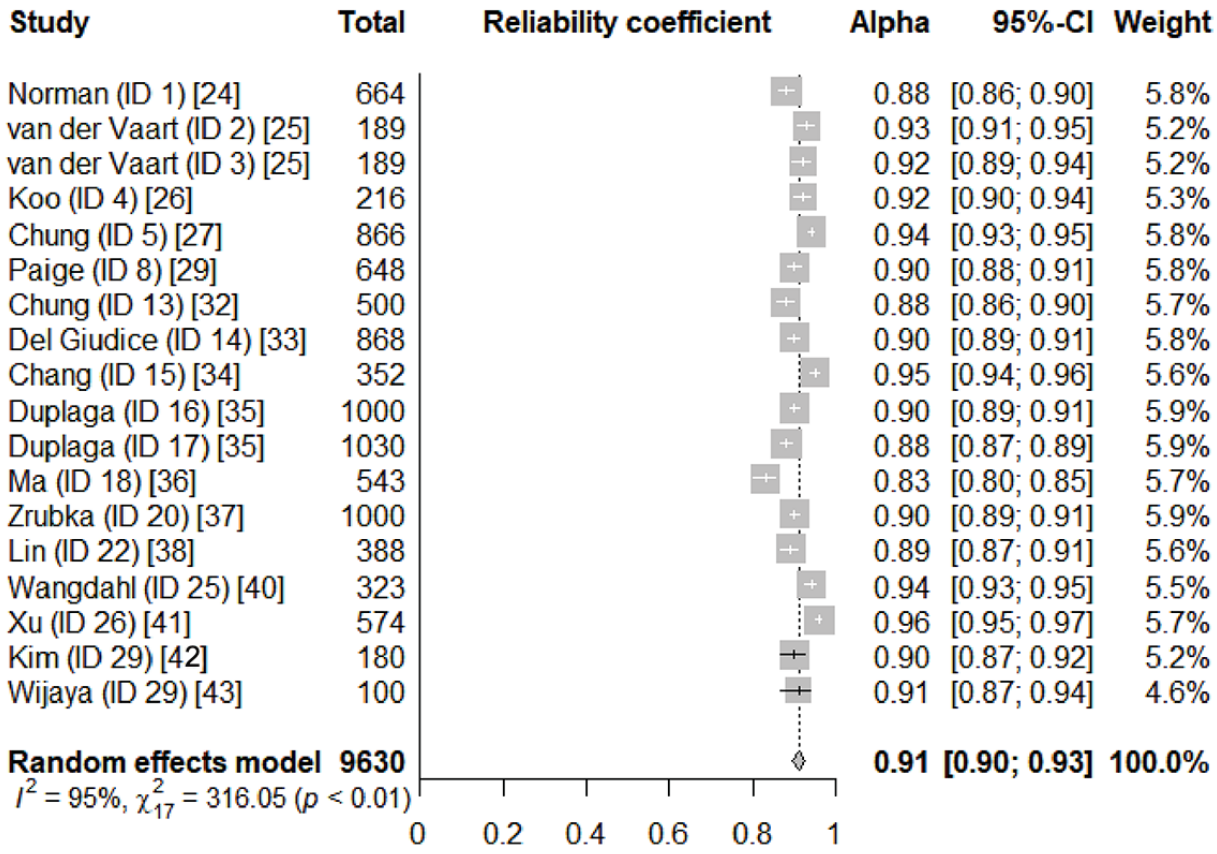
[r]DHLI: digital health literacy instrument.

[s]eHLA: eHealth literacy assessment toolkit.

[t]eHLQ: eHealth literacy questionnaire.

[u]TeHLI: transactional eHealth literacy instrument.

Figure 2. Forest plot of the Cronbach's alphas for the eight-item single-factor eHEALS. eHEALS: eHealth literacy scale; ID: study identification number.



The DHLI, eHLQ, and TeHLI were each psychometrically evaluated twice. Regarding the DHLI, a seven-factor structure (ID 51) [61] yielded sufficient low-quality evidence for structural validity. The high-quality evidence for internal consistency was downgraded to low-quality evidence based on the low-quality evidence for structural validity. There was sufficient low-quality evidence for reliability and insufficient high-quality evidence for convergent validity. The five-factor structure of the DHLI (ID 52) [42] also demonstrated sufficient low-quality evidence for structural validity. The eHLQ from 2 studies reported in a single article (IDs 54, 55) [63] had insufficient high-quality evidence for structural validity, sufficient high-quality evidence for internal consistency, and indeterminate low-quality evidence for measurement invariance. The TeHLI from 2 studies in a single article (IDs 56, 57) [64] demonstrated sufficient high-quality evidence for both structural validity and internal consistency and inconsistent low-quality evidence for convergent validity.

The remaining instruments were assessed only once. The eHEALS-E demonstrated sufficient high-quality evidence for both structural validity and internal consistency and insufficient low-quality evidence for known-groups validity (ID 49) [59]. The e-HLS showed insufficient low-quality evidence for structural validity and insufficient very low-quality evidence for convergent validity (ID 50) [60]. The eHLA exhibited indeterminate, very low-quality evidence for structural validity and insufficient very low-quality evidence for internal consistency (ID 53) [62].

Discussion

Principal Findings

This systematic review found that 7 eHealth literacy instruments are currently available. The measurement properties were most frequently assessed for the eHEALS in 18 languages, 26 countries, and diverse populations (eg, patients, adolescents, adults, and the elderly). The conceptualization of a construct to be measured is a basic and initial step when developing a self-reported instrument. The eHEALS was developed based on the definition of "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem (p. 2)" from the 6 components of literacy in the Lily model: traditional, information, media, health, computer, and science literacies [10]. However, this definition was based on the first generation of simple health information technology (Web 1.0), which later resulted in the eHEALS being criticized as not being sufficiently comprehensive to measure the skills needed for the dynamic and social nature of eHealth (Web 2.0) [68]. One of the researchers who developed the instrument also noticed the lack of social media–related skills being included in the

eHEALS, and suggested updating the instrument [69]. In other words, the eHEALS measures eHealth literacy within the restricted scope of the environment before the rise of social media and the mobile web.

Content validity refers to the degree to which the content of an instrument adequately reflects the construct to be measured [70]. The overall ratings of the e-HLS, DHLI, eHLA, eHLQ, and TeHLI for content validity were high but graded as low-quality to very low-quality evidence for relevance, comprehensiveness, and comprehensibility. These findings imply weakness in terms of whether the high ratings are trustworthy and make it difficult to determine which of the instruments has superior content validity.

The eHEALS is a pioneering instrument measuring eHealth literacy and was originally developed with a single-factor structure. However, various other factor structures were identified in this study: 7 types of two-factor structures, 3 types of three-factor structures, and 1 bifactor structure. A possible reason for such diverse factor structures is the instrument contents when considering that insufficient content validity can impair structural validity [18]. The theoretical basis of the eHEALS was the Lily model, which explained multiple components of the constructs. If the contents of the eHEALS effectively reflected the model, the instrument would have been multidimensional. Item variability was also questioned, even though the eHEALS had the same factor structure. These item inconsistencies (Table 3) might be caused by cultural differences that could be closely related to the digital environment of the country in which the study was conducted. The inconsistencies of the factor structures and the corresponding items might also be due to eHEALS already being outdated for use in evaluations at this time, which reflects the dynamic and social nature of eHealth. It was noticed that the eHEALS items do not assess interactive skills when utilizing the internet [61]. Similarly, this systematic review found inconsistent low-quality evidence for relevance and insufficient very low-quality evidence for comprehensiveness in the eHEALS, which might explain its incongruent structures.

The three-factor eHEALS structures reported on in a single study by Paige et al [58] and 3 studies by Sudbury-Riley et al [54], Gartrell et al [55], and Brørs et al [56] were found to be the best structures, with sufficient high-quality evidence for structural validity, internal consistency, measurement invariance, and known-groups validity. Despite these good measurement properties, the three-factor eHEALS structure reported by Sudbury-Riley et al [54], Gartrell et al [55], and Brørs et al [56] demonstrated insufficient low-quality evidence for reliability and insufficient high-quality evidence for hypothesis testing. The three-factor eHEALS by Paige et al [58] has been evaluated only once, so current evidence of its quality is based on the results of the single study (some measurement properties were not evaluated; thus, no evidence existed). Further study is strongly recommended for the suggested three-factor structures of the eHEALS, including reliability, convergent validity, and responsiveness tests. In addition, the three-factor eHEALS has a lack of conceptual comprehensiveness of eHealth literacy. Revising or updating the contents of the eHEALS is therefore recommended to reflect better the skills required for the social nature of eHealth (eg, the sharing of health information).

The eHEALS-E is the extended version of the eHEALS developed to cover better the complex factors contributing to eHealth literacy. However, that instrument was extended under the same definition used for the original version in 2006 [24]. Therefore, this extended version may also be designated as an instrument measuring a narrow scope of eHealth literacy, as for the eHEALS.

Along with the evolution of interactive communication technologies on the internet, conceptual extensions have been demanded for eHealth literacy. This has resulted in the development of second-generation instruments (eg, e-HLS, DHLI, eHLA, eHLQ, and TeHLI) to measure a wider range of eHealth literacy concepts to make them more suitable for people living in the social-media era of eHealth. However, those instruments have been assessed only once or twice, with there being little meaningful synthesized evidence for the measurement properties of each instrument; therefore, further psychometrics studies of them are strongly recommended.

The TeHLI seems to be psychometrically better than the other second-generation instruments. In addition, this is a theory-driven instrument derived from the TMeHL [66] and based on th**e** measurement of transactional features afforded by online media. However, this instrument has only been assessed twice in a single study using classical test theory and item response theory (IRT)/Rasch model with a specific population (ie, baby boomer and older adult patients with chronic lung disease). Therefore, its synthesized evidence for measurement properties cannot be generalized to healthy people or patients of different ages with other diseases. It is therefore suggested that this instrument needs to be assessed in other populations.

Implications for Future Studies on eHealth Literacy Instruments

The measurement error and responsiveness were not assessed for any of the instruments identified in this study, so future studies of those properties are warranted. More studies of measurement properties also need to be conducted for the second-generation instruments that have been assessed only once or twice. Further psychometric evaluations will increase the credibility of the synthesized evidence. When developing a self-reported instrument, specifying the definition of the concept to be measured is the most basic and important starting point because this determines the scope of the instrument being developed and affects its measurement properties. Nevertheless, the definitions of eHealth literacy were not clarified for most of the instruments identified in this study. New instruments need to be developed for which the definition of eHealth literacy to be measured is clearly addressed, particularly encompassing the attributes/skills required for the social nature of eHealth in the current digital environment.

The assumptions of unidimensionality, local independence, and monotonicity underlying the analyses of structural validity performed using the IRT/Rasch model were not or only partially reported for 11 of 14 studies. According to the COSMIN methodology, the structural validity of a study cannot be rated

as sufficient without information about the nonviolation of assumptions underlying IRT/Rasch analysis, even when the model exhibits an adequate fit for structural validity [18,21]. Therefore, clear reports on whether all assumptions are met are needed for future studies that apply IRT/Rasch analysis to assess structural validity.

Convergent validity refers to the relationship of an instrument's score (eg, eHealth literacy instrument) with a comparator instrument that measures similar constructs and has satisfactory measurement properties [71]. The methodological quality of convergent validity was inadequate in 44% of the studies in this review due to no information being provided on the measurement properties of the comparator instrument(s) used for these assessments of eHealth literacy instruments. Future evaluations of convergent validity should therefore employ comparator instruments with satisfactory psychometric properties.

Regarding the instructions provided for how to respond to items, only those for the eHEALS included the recall period: "…tell me which responses best reflect your opinion and experience right now." Other instruments did not provide information about the recall period, which may result in bias in response items. In the future, it is recommended to provide information about the item response time frame, such as a "short" recall period or the "current state" [72].

Strengths and Limitations

The first strength of this systematic review is that a two-phase search strategy was performed to exhaustively identify eHealth literacy instruments, as recommended by Lee et al [14], especially when searching for concepts involving compound words such as "eHealth" literacy. The second strength is that internal consistency (Cronbach's α) was qualitatively summarized and quantitatively pooled in a meta-analysis. This is the first meta-analysis applying Cronbach's α to eHealth literacy instruments. A limitation of this study is that it only included peer-reviewed journal articles published in English, which may have resulted in selection bias.

Conclusions

This systematic review identified 7 eHealth literacy instruments, and complete evaluations of all measurement properties have not been performed for any of these instruments. The eHEALS, based on the 6 components of literacy in the Lily model, was the most frequently investigated instrument with the smallest number of items (8 items), and the 2 three-factor structures of the eHEALS were better than other structures of the instrument; however, this instrument measures a narrow scope of eHealth literacy and so needs to be reconsidered when being applied to people living in the social media era of eHealth (web 2.0). Revising or updating the contents of the eHEALS is necessary to reflect the skills required for the social nature of eHealth. The TeHLI (consisting of 18 items) was the best instrument for broader measurements of eHealth literacy, although it is restricted by generalizing for only healthy people or patients with other diseases in different ages (younger than 40 years). Further psychometric studies of the second-generation eHealth literacy instruments are strongly recommended. In particular, their content validities should be carefully considered due to the results of this systematic review indicating it had low-quality or very low-quality evidence, meaning that they do not fully capture eHealth literacy.

Acknowledgments

This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT; NRF-2021R1A2B5B01001603). The funder did not play any role in the conduct or publication of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Searching filters.
[DOCX File , 16 KB - jmir_v23i11e30644_app1.docx]

Multimedia Appendix 2
Characteristics of included instruments and studies.
[DOCX File , 44 KB - jmir_v23i11e30644_app2.docx]

Multimedia Appendix 3
Summary or pooled results for measurement properties for each instrument.
[DOCX File , 23 KB - jmir_v23i11e30644_app3.docx]

References

1. Lee E, Lee YW, Moon SH. A structural equation model linking health literacy to self-efficacy, self-care activities, and health-related quality of life in patients with type 2 diabetes. Asian Nurs Res 2016 Mar;10(1):82-87 [FREE Full text] [doi: 10.1016/j.anr.2016.01.005] [Medline: 27021840]

2. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007 Oct;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
3. Zhang NJ, Terry A, McHorney CA. Impact of health literacy on medication adherence: A systematic review and meta-analysis. Ann Pharmacother 2014 Jun;48(6):741-751. [doi: 10.1177/1060028014526562] [Medline: 24619949]
4. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
5. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patient with long-term conditions: A systematic review. Eur Psychol 2019 ;24(1):81. [doi: 10.1027/1016-9040/a000350]
6. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: A cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
7. Internet World Stats. World internet usage and population statistics 2021 year-Q1 estimates. 2021. URL: https://www.internetworldstats.com/stats.html [accessed 2021-03-16]
8. Ratzan SC. Web 2.0 and health communication. J Health Commun 2011;16 Suppl 1:1-2. [doi: 10.1080/10810730.2011.601967] [Medline: 21843091]
9. Fox S, Duggan M. Health Online 2013. Pew Research Internet Project. 2013. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2021-03-18]
10. 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]
11. Horigan G, Davies M, Findlay-White F, Chaney D, Coates V. Reasons why patients referred to diabetes education programmes choose not to attend: A systematic review. Diabet Med 2017 Jan;34(1):14-26. [doi: 10.1111/dme.13120] [Medline: 26996982]
12. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
13. Department of Health and Human Services. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006 Oct 11. URL: http://www.fda.gov/downloads/Drugs [accessed 2021-08-18]
14. Lee J, Lee E, Chae D. Self-efficacy instruments for type 2 diabetes self-care: A systematic review of measurement properties. J Adv Nurs 2020 May 13:2046-2059. [doi: 10.1111/jan.14411] [Medline: 32400902]
15. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowl Manag Elearn 2015 Dec;7(4):576-600 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.038]
16. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
17. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
18. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A delphi study. Qual Life Res 2018 May;27(5):1159-1170 [FREE Full text] [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]
19. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Oct;18(8):1115-1123 [FREE Full text] [doi: 10.1007/s11136-009-9528-5] [Medline: 19711195]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
21. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials 2016 Sep 13;17(1):449 [FREE Full text] [doi: 10.1186/s13063-016-1555-2] [Medline: 27618914]
22. Linacre J. Winsteps® Rash Measurement Computer Program User's Guide. Beaverton: winsteps.com; 2019.
23. Lee J, Lee E, Kim C, Moon SH. Diabetes-related emotional distress instruments: A systematic review of measurement properties. Int J Nurs Stud 2015 Dec;52(12):1868-1878. [doi: 10.1016/j.ijnurstu.2015.07.004] [Medline: 26271434]
24. Norman CD, Skinner HA. eHEALS: The eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
25. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth literacy scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
26. Koo M, Norman C, Hsiao-Mei C. Psychometric evaluation of a Chinese version of the eHealth literacy scale (eHEALS) in school age children. Int Electron J Health Educ 2012;15(1):29-36.

27. Chung S, Nahm E. Testing reliability and validity of the eHealth literacy scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
28. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct validity of the eHealth literacy scale (eHEALS) among two adult populations: A rasch analysis. JMIR Public Health Surveill 2016 May 20;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
29. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
30. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: Cross-sectional survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
31. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth literacy scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
32. Chung S, Park BK, Nahm ES. The Korean eHealth literacy scale (K-eHEALS): Reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
33. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation between eHealth literacy and health literacy using the eHealth literacy scale and real-life experiences in the health sector as a proxy measure of functional health literacy: Cross-sectional web-based survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
34. Chang A, Schulz PJ. The measurements and an elaborated understanding of Chinese eHealth literacy (C-eHEALS) in chronic patients in china. Int J Environ Res Public Health 2018 Jul 23;15(7) [FREE Full text] [doi: 10.3390/ijerph15071553] [Medline: 30041420]
35. Duplaga M, Sobecka K, Wójcik S. The reliability and validity of the telephone-based and online polish eHealth literacy scale based on two nationally representative samples. Int J Environ Res Public Health 2019 Sep 03;16(17):3216 [FREE Full text] [doi: 10.3390/ijerph16173216] [Medline: 31484338]
36. Ma Z, Wu M. The psychometric properties of the Chinese eHealth literacy scale (C-eHEALS) in a Chinese rural population: Cross-sectional validation study. J Med Internet Res 2019 Oct 22;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
37. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth literacy scale. Eur J Health Econ 2019 Jun;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
38. Lin C, Broström A, Griffiths MD, Pakpour AH. Psychometric evaluation of the Persian eHealth literacy scale (eHEALS) among elder Iranians with heart failure. Eval Health Prof 2020 Dec;43(4):222-229. [doi: 10.1177/0163278719827997] [Medline: 30744419]
39. Juvalta S, Kerry MJ, Jaks R, Baumann I, Dratva J. Electronic health literacy in Swiss-German parents: Cross-sectional study of eHealth literacy scale unidimensionality. J Med Internet Res 2020 Mar 13;22(3):e14492 [FREE Full text] [doi: 10.2196/14492] [Medline: 32167476]
40. Wångdahl J, Jaensson M, Dahlberg K, Nilsson U. The Swedish version of the electronic health literacy scale: Prospective psychometric evaluation study including thresholds levels. JMIR Mhealth Uhealth 2020 Feb 24;8(2):e16316 [FREE Full text] [doi: 10.2196/16316] [Medline: 32130168]
41. Xu RH, Zhou L, Lu SY, Wong EL, Chang J, Wang D. Psychometric validation and cultural adaptation of the simplified Chinese eHealth literacy scale: Cross-sectional study. J Med Internet Res 2020 Dec 07;22(12):e18613 [FREE Full text] [doi: 10.2196/18613] [Medline: 33284123]
42. Kim H, Yang E, Ryu H, Kim HJ, Jang SJ, Chang SJ. Psychometric comparisons of measures of eHealth literacy using a sample of Korean older adults. Int J Older People Nurs 2021 May;16(3):e12369. [doi: 10.1111/opn.12369] [Medline: 33527701]
43. Wijaya MC, Kloping YP. Validity and reliability testing of the Indonesian version of the eHealth literacy scale during the COVID-19 pandemic. Health Informatics J 2021;27(1):1460458220975466 [FREE Full text] [doi: 10.1177/1460458220975466] [Medline: 33446030]
44. Soellner R, Huber S, Refer M. The concept of eHealth literacy and its measurement: German translation of the eHEALS. J Media Psychol 2014 ;26(1):38. [doi: 10.1027/1864-1105/a000104]
45. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Health Psychol 2015;17(6):280.
46. Bazm S, Mirzaei M, Fallahzadeh H, Bazm R. Validity and reliability of the Iranian version of eHealth literacy scale. J Community Health Res 2016;5(2):121-130.

47. Richtering SS, Morris R, Soh SE, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: rasch analyses. PLoS One 2017;12(4):e0175372. [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
48. Efthymiou A, Middleton N, Charalambous A, Papastavrou E. Adapting the eHealth literacy scale for carers of people with chronic diseases (eHeals-Carer) in a sample of Greek and Cypriot carers of people with dementia: Reliability and validation study. J Med Internet Res 2019 Nov 28;21(11):e12504 [FREE Full text] [doi: 10.2196/12504] [Medline: 31778120]
49. Dale JG, Lüthi A, Fundingsland Skaraas B, Rundereim T, Dale B. Testing measurement properties of the Norwegian version of electronic health literacy scale (eHEALS) in a group of day surgery patients. J Multidiscip Healthc 2020;13:241-247 [FREE Full text] [doi: 10.2147/JMDH.S242985] [Medline: 32210568]
50. Shiferaw KB. Validation of the Ethiopian version of eHealth literacy scale (ET-eHEALS) in a population with chronic disease. Risk Manag Healthc Policy 2020;13:465-471 [FREE Full text] [doi: 10.2147/RMHP.S240829] [Medline: 32547277]
51. Gazibara T, Cakic J, Cakic M, Pekmezovic T, Grgurevic A. eHealth and adolescents in Serbia: Psychometric properties of eHeals questionnaire and contributing factors to better online health literacy. Health Promot Int 2018 May 25:770-778. [doi: 10.1093/heapro/day028] [Medline: 29800141]
52. Tomás C, Queirós P, Ferreira T. Análise das propriedades psicométricas da versão portuguesa de um instrumento de avaliação de e-Literacia em Saúde. Rev Enf Ref 2014 Jun 30;IV Série(Nº 2):19-28. [doi: 10.12707/RIV14004]
53. Holch P, Marwood JR. eHealth literacy in UK teenagers and young adults: Exploration of predictors and factor structure of the eHealth literacy scale (eHEALS). JMIR Form Res 2020 Sep 08;4(9):e14450 [FREE Full text] [doi: 10.2196/14450] [Medline: 32897230]
54. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth literacy scale (eHEALS) among baby boomers: A multinational test of measurement invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
55. Gartrell K, Han K, Trinkoff A, Cho H. Three-factor structure of the eHealth literacy scale and its relationship with nurses' health-promoting behaviours and performance quality. J Adv Nurs 2020 Oct;76(10):2522-2530. [doi: 10.1111/jan.14490] [Medline: 33463741]
56. Brørs G, Wentzel-Larsen T, Dalen H, Hansen TB, Norman CD, Wahl A, CONCARD Investigators. Psychometric properties of the Norwegian version of the electronic health literacy scale (eHEALS) among patients after percutaneous coronary intervention: Cross-sectional validation study. J Med Internet Res 2020 Jul 28;22(7):e17312 [FREE Full text] [doi: 10.2196/17312] [Medline: 32720900]
57. Hyde LL, Boyes AW, Evans TJ, Mackenzie LJ, Sanson-Fisher R. Three-factor structure of the eHealth literacy scale among magnetic resonance imaging and computed tomography outpatients: A confirmatory factor analysis. JMIR Hum Factors 2018 Mar 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
58. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: Measurement invariance study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
59. Petrič G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 04;19(10):e331 [FREE Full text] [doi: 10.2196/jmir.7372] [Medline: 28978496]
60. 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]
61. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: Measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
62. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: Development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
63. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: Development and initial validity testing of the eHealth literacy questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
64. Paige SR, Stellefson M, Krieger JL, Miller MD, Cheong J, Anderson-Lewis C. Transactional eHealth literacy: Developing and testing a multi-dimensional instrument. J Health Commun 2019;24(10):737-748 [FREE Full text] [doi: 10.1080/10810730.2019.1666940] [Medline: 31583963]
65. Bandura A. Self-efficacy: The Exercise of Control. New York: W. H. Freeman; 1997.
66. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: Concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
67. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag Elearn 2015 Dec;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]

68. van der Vaart R, Drossaert CHC, de Heus M, Taal E, van de Laar MAFJ. Measuring actual eHealth literacy among patients with rheumatic diseases: A qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
69. Norman C. eHealth literacy 2.0: Problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
70. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
71. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: Practical Guides to Biostatistics and Epidemiology. London: Cambridge University Press; 2011.
72. Polit DF, Yang FM. Measurement and the Measurement of Change. Wolters Kluwer: Wolters Kluwer; 2016.

Abbreviations

COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments
DHLI: digital health literacy instrument
eHEALS: eHealth literacy scale
eHEALS-E: eHealth literacy scale-extended
eHLA: eHealth literacy assessment toolkit
eHLF: eHealth literacy framework
eHLQ: eHealth literacy questionnaire
e-HLS: electronic health literacy scale
GRADE: grading of recommendations assessment, development, and evaluation
IRT: item response theory
PRISMA: preferred reporting items for systematic reviews and meta-analyses
TeHLI: transactional eHealth literacy instrument
TMeHL: transactional model of eHealth literacy

Edited by R Kukafka, G Eysenbach; submitted 23.05.21; peer-reviewed by S Paige, L Kayser, Z Zrubka, A Dima; comments to author 30.07.21; revised version received 14.09.21; accepted 04.10.21; published 15.11.21.

Please cite as:
Lee J, Lee EH, Chae D
eHealth Literacy Instruments: Systematic Review of Measurement Properties
J Med Internet Res 2021;23(11):e30644
URL: https://www.jmir.org/2021/11/e30644
doi:10.2196/30644
PMID:34779781

Original Paper

Analysis of a Web-Based Dashboard to Support the Use of National Audit Data in Quality Improvement: Realist Evaluation

Natasha Alvarado[1,2], BA, MA, PhD; Lynn McVey[1,2], PhD; Mai Elshehaly[2,3], PhD; Joanne Greenhalgh[4], PhD; Dawn Dowding[5], PhD; Roy Ruddle[6,7], PhD; Chris P Gale[7,8,9], PhD; Mamas Mamas[10], DPhil; Patrick Doherty[11], PhD; Robert West[12], DPhil; Richard Feltbower[7,13], PhD; Rebecca Randell[1,2], PhD

[1]Faculty of Health Studies, University of Bradford, Bradford, United Kingdom
[2]Wolfson Centre for Applied Health Research, Bradford, United Kingdom
[3]Faculty of Engineering and Informatics, University of Bradford, Bradford, United Kingdom
[4]School of Sociology and Social Policy, University of Leeds, Leeds, United Kingdom
[5]School of Health Sciences, University of Manchester, Manchester, United Kingdom
[6]School of Computing, University of Leeds, Leeds, United Kingdom
[7]Leeds Institute for Data Analytics, Leeds, United Kingdom
[8]Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, United Kingdom
[9]Department of Cardiology, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom
[10]Keele Cardiovascular Group, School of Medicine, Keele University, Keele, United Kingdom
[11]Department of Health Sciences, University of York, York, United Kingdom
[12]Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom
[13]School of Medicine, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Natasha Alvarado, BA, MA, PhD
Faculty of Health Studies
University of Bradford
Great Horton Road
Bradford, BD7 1DP
United Kingdom
Phone: 44 07715433565
Email: n.alvarado@bradford.ac.uk

Abstract

Background: Dashboards can support data-driven quality improvements in health care. They visualize data in ways intended to ease cognitive load and support data comprehension, but how they are best integrated into working practices needs further investigation.

Objective: This paper reports the findings of a realist evaluation of a web-based quality dashboard (QualDash) developed to support the use of national audit data in quality improvement.

Methods: QualDash was co-designed with data users and installed in 8 clinical services (3 pediatric intensive care units and 5 cardiology services) across 5 health care organizations (sites A-E) in England between July and December 2019. Champions were identified to support adoption. Data to evaluate QualDash were collected between July 2019 and August 2021 and consisted of 148.5 hours of observations including hospital wards and clinical governance meetings, log files that captured the extent of use of QualDash over 12 months, and a questionnaire designed to assess the dashboard's perceived usefulness and ease of use. Guided by the principles of realist evaluation, data were analyzed to understand how, why, and in what circumstances QualDash supported the use of national audit data in quality improvement.

Results: The observations revealed that variation across sites in the amount and type of resources available to support data use, alongside staff interactions with QualDash, shaped its use and impact. Sites resourced with skilled audit support staff and established reporting systems (sites A and C) continued to use existing processes to report data. A number of constraints influenced use of QualDash in these sites including that some dashboard metrics were not configured in line with user expectations and staff were not fully aware how QualDash could be used to facilitate their work. In less well-resourced services, QualDash automated parts of their reporting process, streamlining the work of audit support staff (site B), and, in some cases, highlighted issues with data

completeness that the service worked to address (site E). Questionnaire responses received from 23 participants indicated that QualDash was perceived as useful and easy to use despite its variable use in practice.

Conclusions: Web-based dashboards have the potential to support data-driven improvement, providing access to visualizations that can help users address key questions about care quality. Findings from this study point to ways in which dashboard design might be improved to optimize use and impact in different contexts; this includes using data meaningful to stakeholders in the co-design process and actively engaging staff knowledgeable about current data use and routines in the scrutiny of the dashboard metrics and functions. In addition, consideration should be given to the processes of data collection and upload that underpin the quality of the data visualized and consequently its potential to stimulate quality improvement.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-033208

(J Med Internet Res 2021;23(11):e28854) doi:10.2196/28854

KEYWORDS

data; QualDash; audit; dashboards; support; quality

Introduction

Background

Health care organizations are complex systems with variations in patient care and outcomes observed nationally and internationally [1,2]. Audit and feedback may help reduce variations in care quality by comparing clinical performance against standards and benchmarks to stimulate data-driven improvement [3]. In the National Health Service (NHS) in England, national audit and feedback are part of a well-established quality improvement program that encompasses over 50 clinical specialties and patient groups [4,5]. Audit suppliers centrally collate and manage data from participating services and produce feedback, with national comparators, with the intention of stimulating quality improvement [6]. The mode and frequency of feedback varies between national audits but includes paper and electronic formats [7]. Feedback, however, is not used uniformly by participating services to stimulate quality improvement [4,6]. Reported constraints on the use of feedback include access to data, data timeliness and quality, metric relevance, and limited resources for data analysis and interpretation [7,8].

Dashboards offer the potential to overcome some of the constraints reported in the use of national audit data. They use visualization techniques intended to ease cognitive load and improve data comprehension [9,10]. In health care, a distinction is made between clinical dashboards that display performance at the level of individual clinicians or patients to inform direct patient care and quality dashboards that show performance at the level of a ward or organization to inform service improvement [11,12]. Despite the increasing use of dashboards in health care, including as part of the recent response to COVID-19 [13,14], evidence regarding how they become integrated into work processes to impact practice is limited [5]. The aim of this study, therefore, is to investigate how, why, and to what extent a novel quality dashboard (QualDash) supported the use of national audit data for quality improvement in an English hospital setting.

Dashboard Development

QualDash is a customizable, web-based dashboard that was designed and evaluated using data from 2 national audits: The Myocardial Ischemia National Audit Project (MINAP) and the Pediatric Intensive Care Audit Network (PICANet). MINAP collects data spanning 130 data fields, contributed by all hospitals in England, Wales, and Northern Ireland that admit patients with acute coronary syndromes [15,16], whereas PICANet collects data from hospitals and services that transport critically ill children to pediatric intensive care units (PICUs) in England, Scotland, Wales, Ireland, and Northern Ireland [5,17]. These 2 audits vary in clinical specialty and metrics. For example, PICANet takes place in pediatric intensive care and includes metrics such as:

- Standardized mortality ratio: Ratio between the observed number of deaths and the number of deaths that would be expected, given the severity of patients' illness at admission
- Unplanned extubations—accidental removal of breathing tubes
- Emergency readmissions within 48 hours of discharge

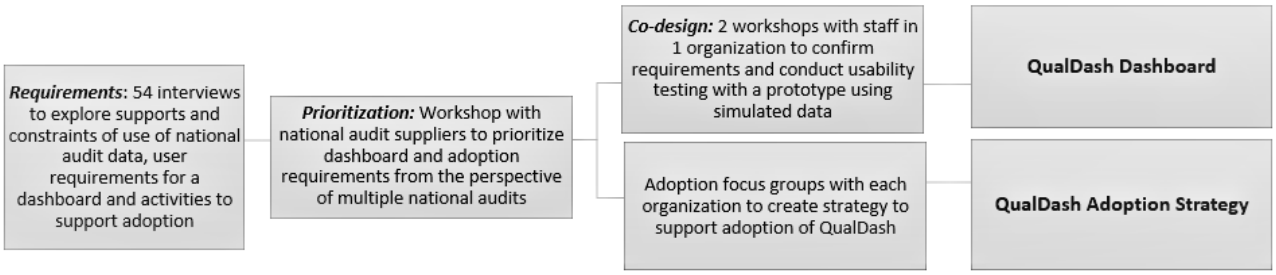
In comparison, MINAP audits services for heart attack and includes metrics such as:

- Call-to-balloon time: Time between ambulance call and primary percutaneous coronary intervention treatment: Target 90 minutes
- Door-to-angiography time: Time between arrival at hospital to diagnostic procedure: Target 72 hours
- Discharge on gold standard drugs: Proportion of patients who received all secondary prevention medication for which they were eligible

These audits were chosen for dashboard development to increase the generalizability of the findings beyond a single audit.

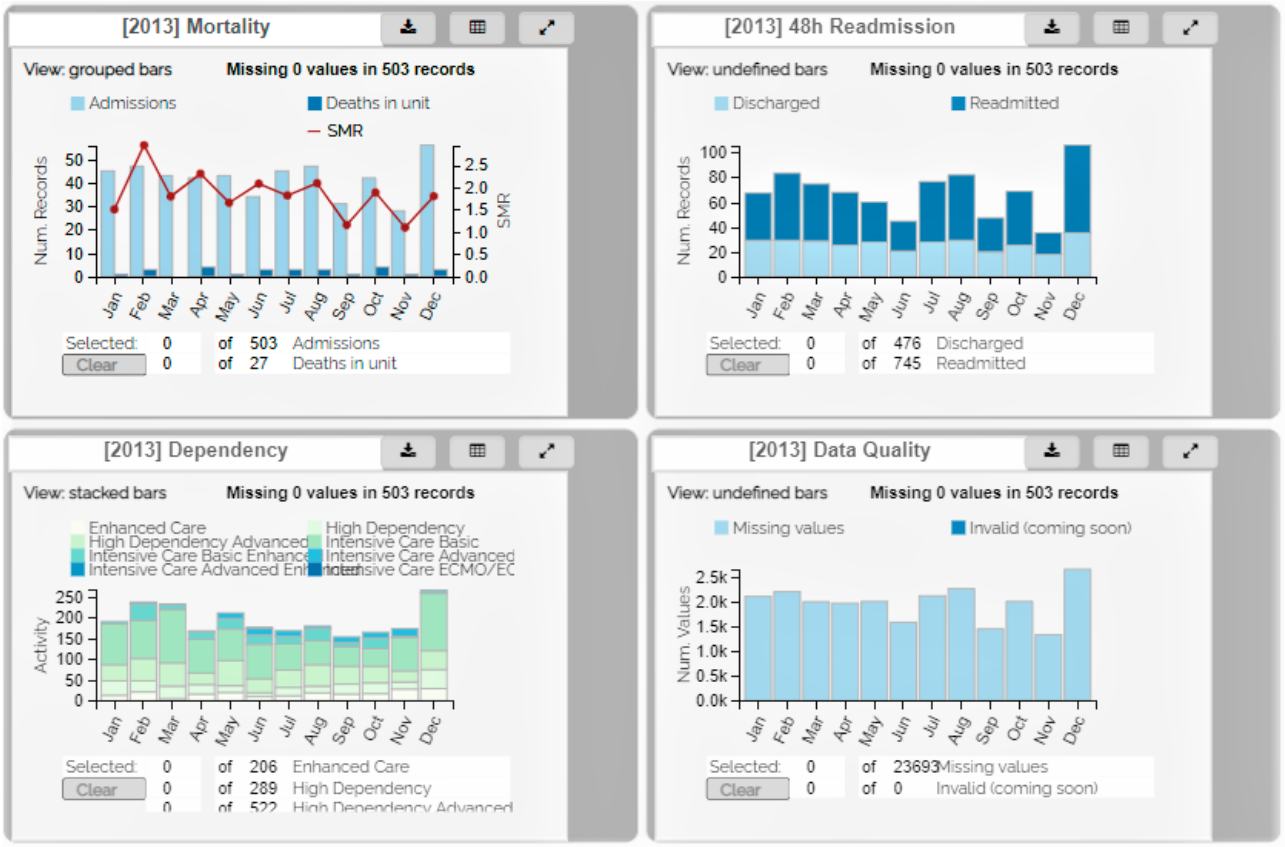
Dashboard development was conducted within 5 NHS acute health care organizations and included interviews with 54 staff members, a workshop with audit suppliers, and 2 co-design workshops with clinicians and managers from one organization [5,7,8,18]. Focus groups were held within each organization to identify strategies to support the uptake and adoption of QualDash. Figure 1 gives an overview of the development work.

Figure 1. QualDash development work.



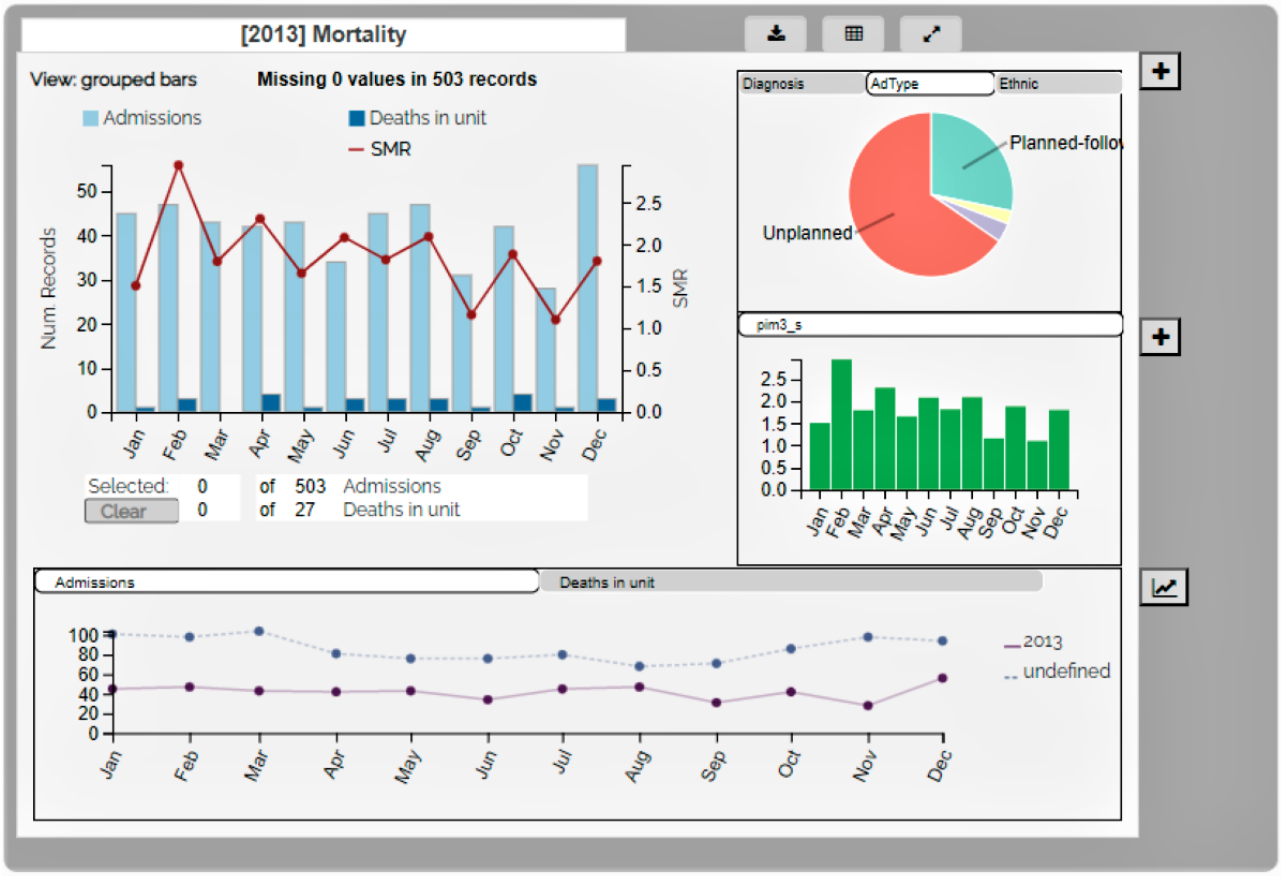
Development work provided insight into how national audit data were used—*user tasks*—and by whom in each organization, and what interrogative and reporting functions a quality dashboard should incorporate to facilitate data use. Requirements were documented in a software requirements specification [11] and were translated into a card metaphor, that is, metrics were configured as bar charts in moveable, customizable areas termed *Qualcards* using a metric specification structure in JavaScript Object Notation. Figure 2 shows an example of the QualDash prototype, displaying 4 Qualcards.

Figure 2. Four Pediatric Intensive Care Audit Network Qualcards—displays simulated data.



Each Qualcard was designed to address a sequence of user tasks related to a single metric. On first opening QualDash, the user would see the main visualization for each metric displayed by month over a year; essentially, these could be used to address key care quality questions quickly and easily, for example, referring to the Mortality Qualcard in Figure 2, a user could address the question of how many deaths were reported in a unit last month compared with previous months? Qualcards were expandable (Figure 3) to display additional and historic measures to answer follow-up questions, for example, what was the method of admission for patients who died on a unit in a given month? Thus, QualDash was designed to facilitate care quality monitoring and data interrogation, where needed.

Figure 3. Expanded mortality Qualcard showing method of admission in the subview pie chart—displays simulated data.



QualDash included functions to export visualizations and raw data and incorporated customization features, for example, users could select which variables were displayed in expanded Qualcard subviews, and dashboard authors (local staff with sufficient information technology [IT] experience) were able to configure service-specific variants of a metric, that is, add or adapt existing Qualcards. Usability evaluation of the prototype dashboard (Figures 2 and 3) was undertaken by staff from 2 organizations, using a think-aloud protocol and simulated data. Having completed a series of tasks, participants completed the System Usability Scale [19], with results suggesting a high level of usability.

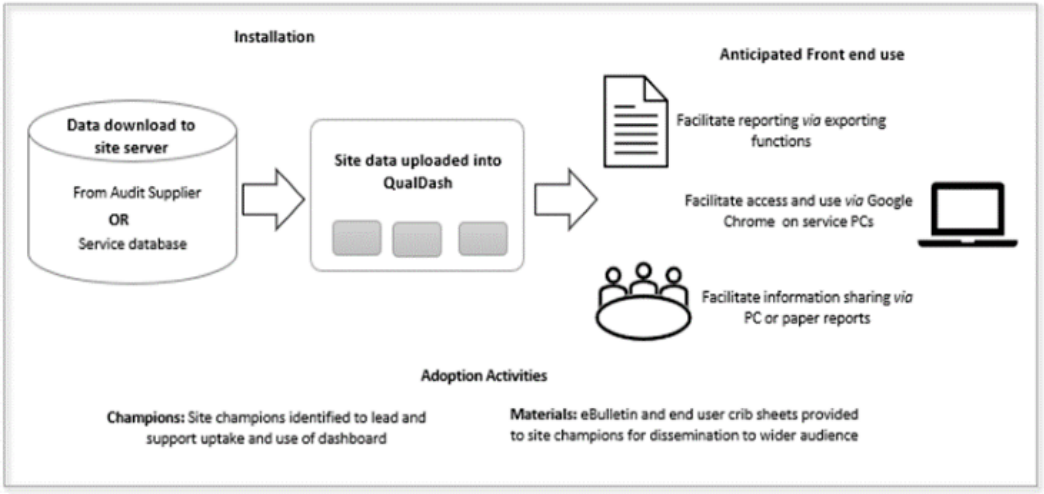
Timeliness of data have also been reported to be important for data use in quality improvement. Managing the dashboard on a central server would have required requesting data from the national audit supplier and then waiting for those data to be sent to the research team before being uploaded to QualDash. Instead, to visualize data that were as timely as possible, QualDash was installed on site servers at each participating organization to enable staff to upload local data as needed. QualDash was accessible via Google Chrome on the service PCs. However, locating QualDash on local servers meant the dashboard displayed site data only; the metrics were relevant and used by services nationally but did not include national comparator data. Therefore, rather than comparing performance against peer organizations, sites compared service performance month-by-month and against national targets, for example, time to treatment targets.

Methods

Study Sample and Dashboard Installation

QualDash was installed in 5 organizations including 3 large teaching hospital trusts that offer specialist services, including PICUs, and 2 smaller district general hospitals. Within these sites, QualDash was evaluated in services that participated in the 2 national audits of interest: PICANet and MINAP. District general hospitals do not provide the PICU service for which PICANet operates; therefore, the PICANet dashboard was available within PICUs in the 3 teaching hospital trusts and the MINAP dashboard was available in all 5 organizations. Figure 4 depicts how service-specific, real data were uploaded to QualDash on installation and anticipated front-end use.

Figure 4. QualDash—data flow and anticipated front-end use.



To promote confidence in the accuracy of the data displayed, site data were validated by the research team on upload, after which site staff were able to upload data at their convenience. To support the adoption of QualDash and informed by the adoption focus groups, the research team identified *champions*, typically the study collaborators at each site, to lead, and support others in the use of the dashboard. Champions were provided with an eBulletin to advertise QualDash and crib sheets, with instructions on how to use dashboard functions, to disseminate as they saw appropriate within their organization. Furthermore, champions were encouraged to provide feedback about the dashboard that could be used to improve functionality where possible.

Evaluation Framework

QualDash was conceptualized as a sociotechnical intervention [20-22]. Therefore, we sought to understand how interactions between the technological (the dashboard itself and existing information technology [IT] systems) and social (individuals and groups) system elements shaped dashboard use and impacts. For this purpose, we used realist evaluation, a theory-driven approach [23]. Realist evaluation is appropriate for evaluating sociotechnical interventions because it acknowledges that intervention impacts vary according to participants' responses to the intervention, which are highly dependent on context. It provides a way to understand such variations using the concepts of mechanism, context, and outcome. Mechanisms refer to how recipients respond to, reason about, and interact with intervention resources, whereas context refers to factors that influence how mechanisms operate, including personal attitudes, beliefs, cultures, and resources [24]. Outcomes refer to the intended and unintended or unexpected impacts of interactions between mechanisms and contexts. Evaluation by this mode involves a cycle of constructing, testing, and refining realist theories configured as Context + Mechanism = Outcome, that is, hypotheses of how, why, and in what circumstances an intervention might work [23].

Multimedia Appendix 1 shows 2 Context + Mechanism = Outcome configurations (CMOcs) constructed using data collected as part of dashboard development [5,7,8], which will be used to illustrate the main study findings. In summary, we understood through development work that there was variable use of national audit data before QualDash, largely because of differences in the resources allocated to data collection and management across sites. However, clinicians used and acted on data if it helped them monitor whether their service was delivering safe and effective care, a mechanism we termed *professionalism*. We expected that QualDash, underpinned by this mechanism, would be used to facilitate data reporting (CMOc 1) or to integrate national audit data into routine reporting (CMOc 2) depending on existing use of data and resources, as illustrated in Multimedia Appendix 1.

Extent of Use of QualDash

To assess the extent of use of QualDash across organizations and services, log files automatically recorded information about the use of QualDash at each site from the time of installation (June and July 2019) to July 2020. The information captured included the type of user (job title), data used (audit, year), time spent interacting with different Qualcards, and functionality used. Data from the log files were postprocessed in Microsoft Excel, removing entries generated by members of the research team undertaking on-site testing of the software. Using the timestamps for each login, the number of sessions per audit per month from installation to the end of July 2020 was determined for each site. Where a login occurred less than 20 minutes after the last timestamp and appeared to be the same user (based on the audit and year selected and the job title entered), this was treated as a continuation of the previous session.

Observations of Practice

To understand how, why, and where QualDash was used, observations of practice were conducted at study sites between August 2019 and February 2020, in spaces where QualDash had the potential to be used in the ways expressed in the CMOcs, including:

- Hospital wards and units (eg, admission wards, catheterization laboratories, acute coronary care units, and PICUs)
- Clinical governance and directorate meetings
- Offices used by audit support and clinical staff for data management and administration activities

- Informal interviews relating to the activities observed

An observation schedule was developed to help guide data collection, including prompts to capture (1) service processes and routines for monitoring care quality, (2) types of data and technologies used and by whom, and (3) how, why, and the extent to which QualDash was integrated within these routines. Local collaborators within each service facilitated observations by introducing researchers to ward matrons and senior nurses. These staff provided permission for the observations to take place and notified their colleagues. Observations were conducted by 2 researchers who recorded the observations in handwritten notes. The observations were nonparticipatory, but researchers interacted with participants via informal interviews where an explanation of the activity observed was provided by the participant. The 2 researchers initially conducted observations together to develop a shared understanding of how to record fieldnotes, after which each researcher aimed to conduct at least one observation per service. Table 1 shows the total number of hours and types of observations conducted at each site.

Table 1. Hours and type of observation performed in study sites.

	Teaching Hospital Trust (MINAP[a] and PICANet[b])			District General Hospital (MINAP)		Total hours
	A	B	C	D[c]	E	
QualDash installation and customization meetings	9	8	16	7	6.5	46.5
Ward and "back office" observations, including data collection and validation	25	24.5	17	2	13.5	82
Meeting observations and informal interviews	7.5	3	4	3	2.5	20
Total hours of observation	41.5	35.5	37	12	22.5	148.5

[a]MINAP: Myocardial Ischemia National Audit Project.

[b]PICANet: Pediatric Intensive Care Audit Network.

[c]QualDash installed in December 2019 because of delays in site approval. This explains why the first use of QualDash at this site, discussed in results, begins in January 2021, much later than at other sites.

Observation notes were transcribed by researchers as soon as possible after the observation and analyzed following (and adapting) the 5 stages of framework analysis [25-28]. To familiarize themselves with the data (stage 1 of Framework Analysis), the transcripts were read independently by 3 researchers who then met to discuss their interpretations and construct themes to categorize the text (stage 2 of the Framework Analysis). The themes were uploaded into NVivo (QSR International; software for facilitating qualitative data analysis) and systematically applied to all transcripts (stage 3 of framework analysis) by the 2 researchers who conducted the observations, with themes being refined and added as necessary to capture the range of experiences. Instead of *charting* (stage 4 of Framework Analysis) where data are organized in charts of themes, NVivo was used to extract data categorized in each theme; these were then summarized in narrative accounts, for example, of the care quality monitoring processes observed, and the data and technologies used in these processes. To test CMOcs, data relevant to each CMOc (identified in the analysis work conducted) were summarized in a matrix developed in Word that displayed each CMOc construct by service. In this way, the CMOcs were refined by comparing observational data from each site against the hypotheses. Researchers used observational data to support the interpretation of the log file analysis by examining if it suggested an explanation for extent of use, for example, if QualDash was being used because it facilitated the work of audit support staff.

Perceived Use and Usefulness

At the end of the evaluation period, a questionnaire based on the technology acceptance model (TAM) was administered [29]. The TAM consisted of 12 statements, 6 concerning usefulness, and 6 concerning ease of use. Respondents rated each statement on a scale of 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement. In addition to the TAM items, the questionnaire included questions regarding how frequently respondents use national audit data, how frequently they used QualDash during the evaluation period, and how likely they would be to continue using QualDash after the evaluation period. A link to the questionnaire was emailed to 35 participants, who were known to have either used QualDash or seen it demonstrated at the beginning of August 2020. Two reminder emails were sent, and the survey was closed at the beginning of September 2020.

The questionnaire data were analyzed to assess the perceived usefulness of QualDash and its ease of use. Excel was used to produce summary statistics for each TAM statement, calculating the mean and range for all participants and for those who reported having used QualDash, broken down by audit and role. Guided by previous studies, ratings of 3 or higher were considered to indicate a positive response.

Ethical Approval

Ethical approval for this study was provided by the University of Leeds School of Healthcare Research Ethics Committee (approval # HREC16–044). Approval was received from the Health Research Authority and each of the 5 sites.

Results

Extent of Use of QualDash

Figure 5 shows the total number of uses of QualDash per clinical service and by site, and Figure 6 shows the number of uses by

site over the evaluation period. The figures show variations in the extent of use across sites and clinical areas, for example, in site A, there was no use of QualDash within the PICU throughout the evaluation period (there could be no PICU use in sites D and E that did not have PICUs), whereas the greatest use was in the site B PICU and site E Cardiology.

Figure 5. Extent of QualDash use by site and clinical service. PICU: pediatric intensive care unit.

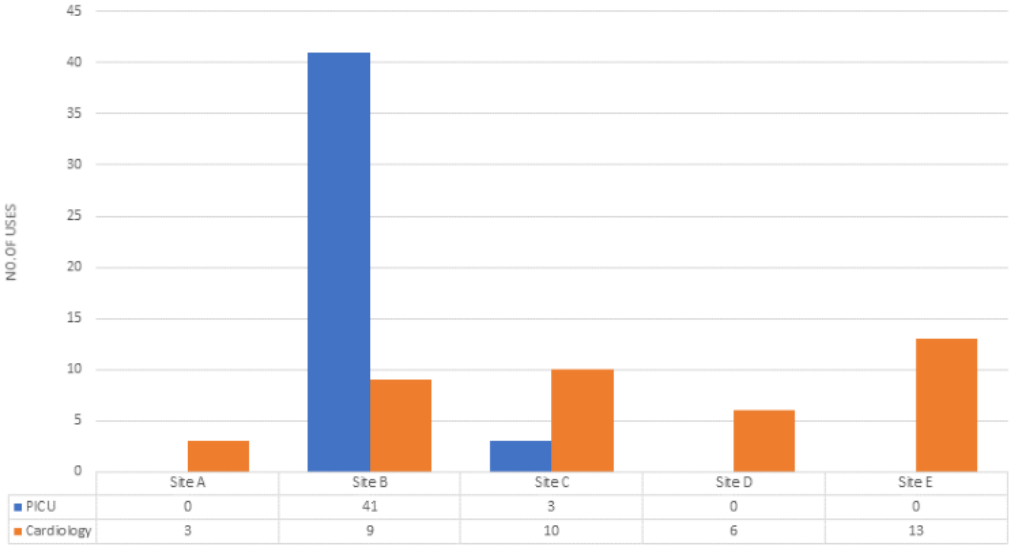
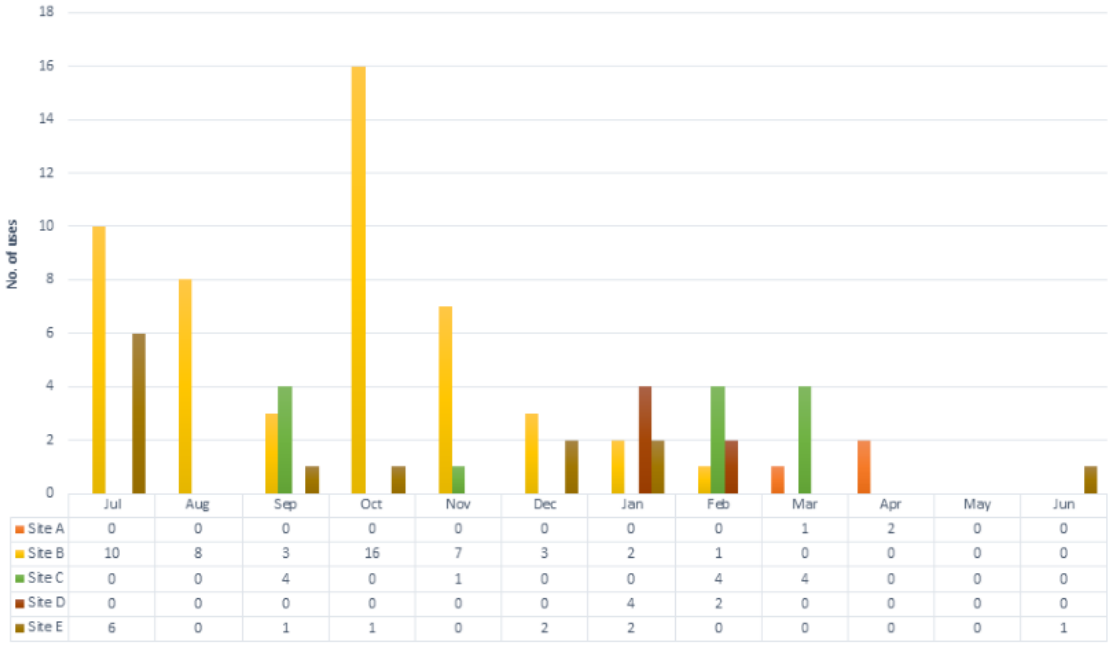


Figure 6. Uses of QualDash by clinical service and month.



Below, we use the observation data to understand the reasons behind these variations and to refine the CMOcs in Multimedia Appendix 1. We begin with a summary of the care quality monitoring routines observed, and then, to refine the CMOcs in Multimedia Appendix 1, explore how, why, and to what extent the dashboard impacted practice by facilitating reporting (CMOc 1) or integrating national audit data in routine practices (CMOc 2).

Observations of Practice

Quality Monitoring Routines

The services observed included PICUs, where clinical staff cared for children, often on ventilators to assist their breathing, and cardiology wards for adults in admission for, and treatment or recovery from, acute coronary syndrome. Quality monitoring processes varied across services but included ward activities directed at the care of individual patients and directorate and clinical governance meetings directed at service-level monitoring of quality and safety.

Ward observations were conducted at or around the nurses' stations. As might be expected, staff routines focused on the needs of individual patients, for example, ward staff shared information about patient care via handovers (summaries at shift changes), ward rounds, and clinical team *huddles* where staff met intermittently throughout the day to update colleagues about each patient's condition and care plan. The use of technologies and data on all wards also focused on individual patient care. For example, electronic whiteboards and dry-erase whiteboards were used to visualize data in matrices, for example, patient by task or activity such as if they were awaiting a discharge letter. The use of patient case notes in paper form or electronic devices, such as PCs, laptops, and iPads positioned around the nurse's station was also frequently observed. QualDash, however, was not observed to be used by staff on wards, despite opportunities via the electronic devices available to them.

The clinical governance (service) and directorate (specialty) meetings were observed monthly or quarterly. These meetings were attended by a range of staff, including nurses, physicians, junior doctors, and service managers. In comparison to the wards, observations revealed that national audit data were used in these meetings at certain sites, sometimes via QualDash. Figure 7 shows the flow of audit data from collection to practice, captured during observations.

Figure 7. Audit data collection, management, and use. MINAP: Myocardial Ischemia National Audit Project; National Clinical Audit; PICANet: Pediatric Intensive Care Audit Network; PICU: pediatric intensive care unit.

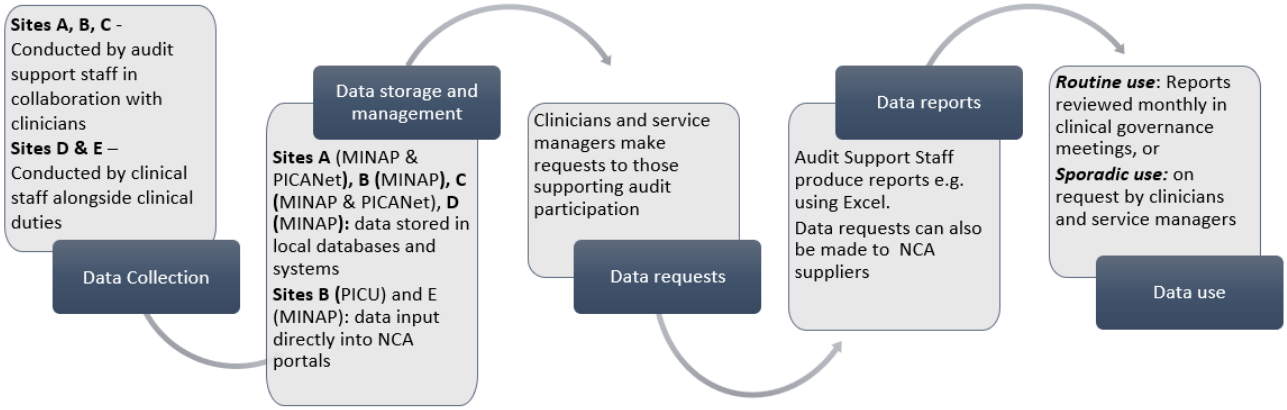


Figure 7 is a simplified model of data use, showing how some services were resourced to use national audit data within their routine monitoring processes, whereas in others, use of national audit data was more sporadic, usually where clinicians were not supported by dedicated audit support staff. Next, we discuss the extent to which QualDash was integrated within work routines, and explore interactions that help explain the extent of use captured in the log files.

Facilitating Use of National Audit Data

The CMOcs hypothesize that, depending on the context, QualDash would be used either to facilitate the use of, or introduce, national audit data into service monitoring processes, such as the clinical governance meetings depicted in Figure 7. In practice, constraints were reported that limited the use of QualDash in these ways.

Sites A and C (PICU and Cardiology units) were resourced with dedicated audit support staff who produced performance reports when requested by clinicians and managers. Reports were produced using local databases and systems such as Microsoft Access or Microsoft Excel, where national audit data were stored before upload to audit suppliers. Observations revealed that, within these services, audit support staff continued to use existing systems post installation of QualDash. In site A Cardiology unit, for example, the nurse dedicated to audit participation did not have Google Chrome installed on their PC, despite repeated requests to their IT department, so they were unable to open QualDash to support reporting requests. Even in services where QualDash was accessible via Google Chrome, however, audit support staff continued to use existing systems, as this observation of an audit clerk updating monthly reports using their access database highlights:

> *We discuss a Qualcard that shows invasive ventilation and bed days. The Audit Clerk checks the figures in Qualcard against her visualizations for the same months, and they do not match. They explain that this mismatch might be because QualDash shows bed days by admission date rather than bed days across the months of admission. In one month, the Qualcard showed that there were more invasive ventilation days than there were bed days (which is impossible). Furthermore, the data in QualDash have not been updated since the installation (in July); therefore, the data in the access database is the only way to update the reports discussed today.* [site A; PICU Observation; September 4, 2019]

The use of simulated data in the co-design workshops meant that staff were unable to scrutinize the dashboard with data familiar and meaningful to them at that time. It was only post installation; therefore, when using QualDash to explore site specific data that it became apparent that some metrics were not configured as typically reported by the service. The issue described above was addressed by incorporating additional functionality within the dashboard, allowing Qualcards to be configured to present information in their respective months. The dashboard customization feature also supported metric reconfiguration. However, QualDash was not able to configure all of the metrics reported locally. For example, the number of accidental extubations (removal of a breathing tube) on PICUs

per month could not be configured in QualDash because of the complex way in which data were captured as part of the audit. The measure was provided as the total number of incidents across the year in the subview pie chart but could not be used to facilitate monthly reporting. Similarly, in the site C Cardiology unit, a MINAP project assistant commented:

> *I would really love to be able to use it [QualDash], but it just does not show me, you know, everything I need, and I would like to be able to say to you, yeah, I use it all the time to add. But [...] it just does not do exactly what I need. But maybe if you spoke to someone like, the lead consultant [physician], and you showed him what was available, and you said, does this answer, you know, some of your queries? However, like the thing he asked me to do, to look at the patients that have failed [within specific time points].* [site C; MINAP Project Assistant, Interview February 2020]

The request to which the MINAP assistant refers was to see the data of patients treated between specific hours in one day. In the work undertaken to capture requirements for QualDash, tasks using such a timeframe were not identified; thus, this functionality was not included in QualDash.

In comparison to the dashboard itself, user knowledge of the audit data set and understanding of QualDash's functionality also constrained use of the dashboard; an audit clerk in site C reported:

> *The Audit Clerk says that the service database contains more detail than PICANet, for example, when they produce reports for the business meeting, [...], they include information about what ward the patients have been referred from and where the patient was discharged. They also get "ad hoc" requests for data on a regular basis.* [site C; Informal interview with PICU Audit Clerk; October 16, 2019]

In fact, QualDash (and the PICANet data set from which it was derived) provided referral and discharge information in Qualcard subviews but this participant was unaware that the dashboard could support this data need. Even so, an additional influence on staff choice to use QualDash in these well-resourced services was that the data displayed were not as timely as those stored in local systems. This was due to challenges in establishing routines for uploading data into QualDash including having the necessary software installed on service PCs to complete the upload. Therefore, despite the manual work involved in report production, audit support staff in these services chose to use existing systems as they provided data that were timely and from which they had expertise that enabled them to configure metrics reported routinely and when requested by clinical team members.

In comparison to sites A and C, the audit clerk in site B PICU consulted hard copy and electronic data sources to submit quarterly reports to NHS England (a body of the Department of Health responsible for commissioning NHS services). Data sources included the ward diary, admissions books, and the patient administration system, and they also consulted with clinicians to retrieve data. QualDash automatically calculated measures, including patient bed days, 48-hour readmission, and mortality rates. Consequently, this audit clerk chose to use QualDash as it facilitated and streamlined their reporting, as hypothesized in the CMOc. Extent of use of QualDash in this site is reflected in Figure 5, and the impact captured in this email sent to the research team: "Just wanted to say QualDash has saved me hours of work with regard to data I submit to NHS England."

An important observation was that, unlike other sites, this audit clerk developed a routine for uploading data into QualDash in a timely way. Therefore, where staff understood and experienced the benefits of using QualDash in their work routines, they supported the data upload process necessary to maintain dashboard use in practice.

Integrating National Audit Data Into Routine Monitoring Processes

Sites D and E were not resourced by staff dedicated to national audit data collection and management. These sites represented an opportunity for QualDash to be used to integrate national audit data within routine monitoring processes. QualDash was observed to be used within these sites, but not in the ways expressed in the CMOcs. For example, in site E, a cardiologist (and champion) was initially keen to use national audit data via QualDash. However, at the postinstallation site visit, the cardiologist noted:

> *The cardiologist says that they have not used QualDash since we first emailed the link to them in the installation email. When asked why, they explain that they have not yet integrated QualDash into their routines. They say that they are not accustomed to having access to MINAP data in this way. Therefore, they sometimes forget that it is there to look at and use QualDash. However, they say that there is no reason why they should not use it, and that one way in which they could try and integrate it into their practice is in their performance meetings where such data could be reviewed.* [site E; informal interview; November 27, 2019]

Despite being involved in the development and dissemination of QualDash as a champion, this physician and their service had managed without routine use of national audit data, and unfamiliarity with access to the data appeared to constrain their interactions with QualDash. The cardiologist added QualDash to the agenda for the next directorate meeting, which was observed as part of the data collection. During the meeting, the functions of QualDash were demonstrated to attendees, including physicians and nurses, and some tentative impacts were highlighted:

> *The cardiologist says that when QualDash was first installed, it enabled them to see where there was missing data and the audit team cleaned up the data in response. Therefore, QualDash has already helped the service to get their data a bit cleaner.* [site E; Cardiology Directorate Meeting; December 11, 2019]

The visualizations in QualDash enabled the audit team to identify missing data and work to address the gaps. However, the meeting chair queried the accuracy of the data displayed:

> *The meeting chair queries the displayed data. They say that all the graphs steadily decline and that "that cannot be accurate." They ask [the Cardiologist who acted as QualDash Champion] why the graphs are like that, who responds that the audit team must be behind with the data collection. The Chair says that it is "amazing" to have access to real-time data, and that they should be able to make more use of MINAP data using QualDash.* [site E; Cardiology Directorate Meeting; December 11, 2019]

The meeting attendees responded positively to QualDash's potential for supporting data use, but the data currently displayed could not be used optimally for quality improvement, because they were incomplete. This finding highlights the importance of the work underpinning front-end use of the dashboard, that is, data collection and upload, and the significance of the dedicated audit support staff in sites better resourced to use national audit data for quality improvement.

In site D, queries about the metric configurations discussed during a postinstallation meeting (QualDash was installed at this site in December 2019) raised concerns about the use and dissemination of the dashboard.

> *[Referring to a configuration of a complicated measure—patients with a certain diagnosis discharged on "gold standard" (up to five) drugs by month and for which the research team has received conflicting definitions] Cardiologists comment that they have come across this sort of thing before with new systems that are developed by people outside their unit: the data going in are correct, but it does not make sense, and that time is needed to check it to ensure that the data are interpreted correctly. They know what their own data should look like, though people outside the unit, including managers from their trust, do not always have that understanding and can misinterpret data that are not displayed correctly.* [site D Observation; January 15, 2020]

On the basis of previous experience with similar initiatives, the champions requested that all Qualcards undergo a *sense check* by staff familiar with site data. The expectation was that this sense check would be completed before dissemination activities took place, due to concerns about data interpretation by staff less familiar with the data. This finding also highlights the role of audit support staff, that their skills and experience in data collection, management, and report production provide confidence in the outputs they create and that site champions would need a similar level of trust in the dashboard to lead its use and dissemination to a wider audience.

Perceived Usefulness and Ease of Use Questionnaire

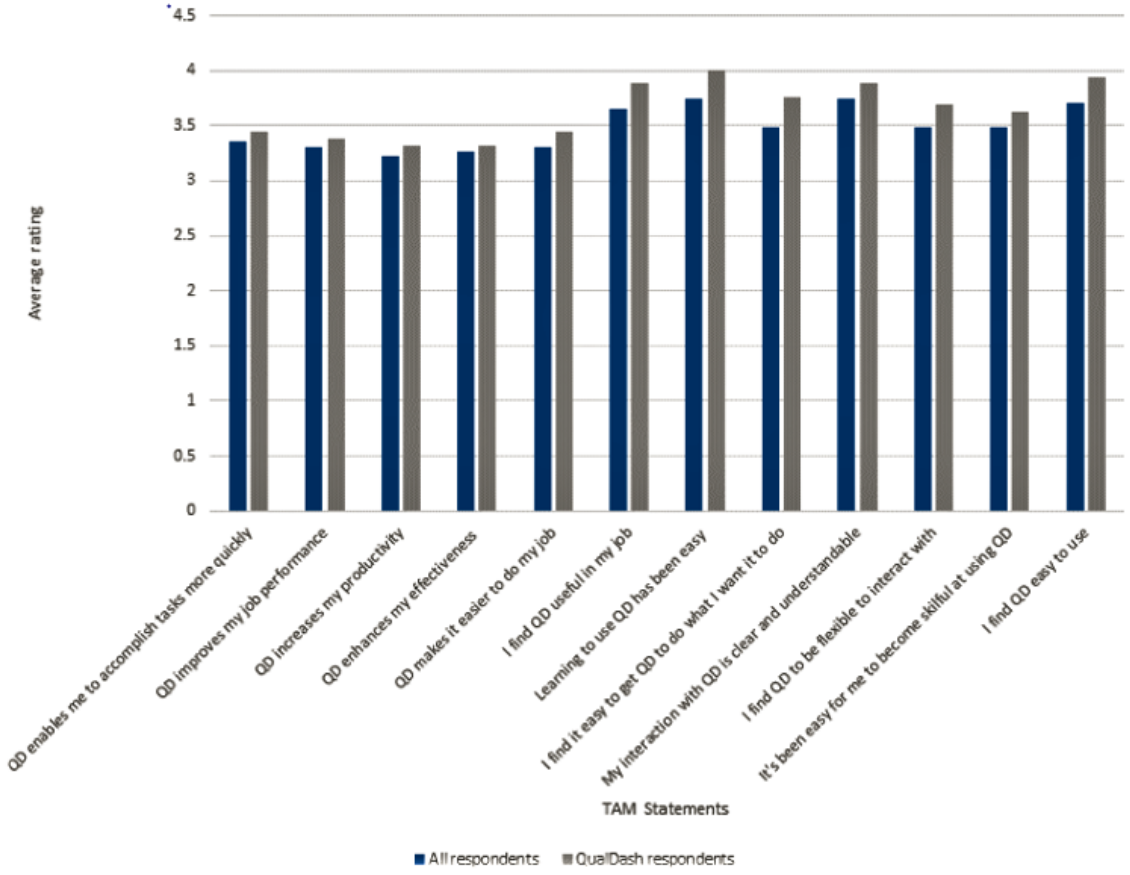
In total, 23 responses were received from the adapted TAM questionnaire, 18 from participants who had experience using QualDash. Figure 8 shows the average ratings of each TAM statement.

Although the average rating for all statements was positive, ratings were higher among those who had used QualDash and higher for ease of use as opposed to usefulness. The questionnaire was anonymous, so we were unable to link responses to sites, but we were able to analyze them according to role, which revealed that audit support staff and doctors found QualDash more useful than nursing staff. On the basis of observations of ward practice, this finding is possibly a reflection of nurses' more frequent engagement with data that informs direct patient care, as opposed to service-level data. Interestingly, however, observations suggested that senior nurses may find QualDash useful, as a pediatrician (and champion) commented:

> *I ask the physician if they see a role for QualDash in the PICU from what we have discussed. The physician says that QualDash is an "amalgamate of retrospective data" but it may help prospective planning, for example, highlighting high-risk patients. They say that they think senior nurses, bands 6 and 7, may be interested in the dashboard, and that QualDash icons on the nurse station PCs might be helpful in supporting their use of the dashboard.* [Pediatrician; informal interview; site A; December 4, 2019]

The pediatrician noted that QualDash displays aggregate, as opposed to patient-specific data, and that this may be useful for prospective planning on the wards. Therefore, senior nurses might utilize the dashboard if desktop icons facilitate access at the nurses' station. Despite constraints on the use of QualDash during the evaluation period, its potential as a useful tool for a range of health care staff was recognized and ideas to support uptake in these ways suggested (desktop icons).

Figure 8. Rating of technology acceptance model statements. QD: QualDash; TAM: technology acceptance model.



Refinement of CMOcs

CMOc 1 in Multimedia Appendix 1 hypothesizes that QualDash would be used to facilitate the use of national audit data where data were used in care quality monitoring. We found that the use of data continued as usual during the evaluation period in services resourced with audit support staff and local systems that supported reporting needs. The reasons behind this choice encompassed dashboard functionality; it was not always possible to configure the Qualcards requested, some of which were used in routine reports—and contextual influences, for example, data were timelier in local systems where they were stored before uploading to QualDash and staff were not always aware of how QualDash could be used to facilitate their work. QualDash was used by an audit support clerk who recognized and experienced benefits (streamlined reporting process) from dashboard use in comparison to their existing system, and consequently established a routine for uploading timely data.

CMOc 2 in Multimedia Appendix 1 hypothesizes that QualDash would be used to integrate national audit data within routine monitoring processes where there were previously constraints accessing data. In one such site, we found that early interactions with QualDash led to additional requests to reconfigure Qualcards in line with user expectations. Work reconfiguring Qualcards was necessary to provide champions, across sites, with confidence to lead use of, and disseminate, QualDash within their service. In another service, we found that the dashboard visualizations highlighted missing data and that staff supporting audit participation could then work to address these issues. However, due to the missing data, QualDash could not be used optimally for quality improvement.

QualDash Version 2 and the Impact of COVID-19

In response to user feedback, a new version of QualDash (Version 2) was developed that addressed issues with Qualcard configurations where possible, including invasive ventilation in PICANet and discharge of gold standard drugs in MINAP, as discussed above. The intention was to install Version 2 across all sites, continue dissemination activities, and capture further impacts via observations. However, in March 2020, the data collection and site visits were suspended due to COVID-19. The lockdown effectively ended face-to-face evaluation work, with data to refine CMOcs focused on learning from QualDash Version 1. Interestingly, the onset of COVID-19 has been accompanied by a rapid uptake in the use of digital technologies to deliver care in lockdown restrictions and to understand the impact of the virus on health care systems [30,31]. In this study, users in a site that previously made limited use of QualDash saw the potential benefits of the dashboard for responding to COVID-19 if it could be adapted to display data on a daily and weekly basis. Therefore, the computer scientist developing QualDash has worked to develop and install a new iteration of QualDash to support services in monitoring the impact of COVID-19.

Discussion

Principal Findings

This paper presents findings from a realist evaluation of a web-based, customizable, quality dashboard designed to support the use of national audit data in quality improvement. QualDash was co-designed with data users to address their needs, and champions were identified to support uptake and adoption. Even so, observations of practice revealed that QualDash had a variable impact across sites within the evaluation period. These findings are comparable with a recently updated review of dashboards that showed variation in impact, even within the same clinical area [32]. We used realist evaluation to understand the reasons behind the variation in impact.

Using CMOcs, we hypothesized that staff would integrate QualDash within their routines because it could facilitate monitoring and interrogation of metrics considered markers of safe and effective care—a mechanism we termed *professionalism*. This mechanism was identified in a context analysis of the use of national audit data and underpinned physicians' use of data in quality improvement [7]. However, observations of practice revealed that when staff interacted with site-specific data via QualDash, they identified that some metrics were not configured in the format they expected. The dashboard's customization feature enabled some metrics to be reconfigured to meet user needs, but it was not possible to configure all metrics requested or reported locally. Therefore, QualDash was not perceived as a tool that could facilitate data use as part of *professionalism* in some sites. Furthermore, not all services were resourced to upload data considered accurate or timely; attributes identified as influencing the use of national audit data and dashboards in health care more generally [6,7,14,33]. These factors constrained the use of QualDash in care quality monitoring, particularly where existing mature systems were in place to support data use.

A study of dashboard design in the context of lymphedema services reported that the complexity and accessibility of data to develop dashboards to support use of aggregate data was more challenging than the development of clinical, patient-level dashboards [33], and a review of data visualization dashboards has highlighted the need for user-centered design and interactions with a prototype to support dashboard development and implementation [34]. Our findings support this work and point to several ways in which the development of quality dashboards could be improved to support uptake and impact.

- First, the use of data familiar and meaningful to users in the co-design process would be beneficial. The use of site-specific data would encourage participants to scrutinize visualizations more thoroughly to confirm that they are configured as required for care quality monitoring before installation. Other studies have highlighted similar messages about how the interpretability of information needs greater consideration in design processes alongside usability and ease of use [35]. Our study highlighted that site champions need to have confidence that metric configurations are aligned with service expectations, to support interpretation by site staff, and to lead its use in quality improvement efforts.
- Second, engaging staff familiar with and knowledgeable about data use within the service, such as audit support staff and physicians, in scrutiny of dashboard metrics and functions, using real data, would provide an opportunity to (1) develop dashboard functionality in ways that can better support existing routines of data use and (2) raise awareness of the range of functionality offered by the dashboard among these key stakeholders, providing them with motivation to integrate the technology into their working practices on installation.
- The findings also emphasize the work surrounding and maintaining dashboard use. Alongside configuring metrics that reflect service needs, consideration needs to be given to the systems in place for collecting and uploading data in a timely and accurate way—attributes that also underpin confidence in the data and consequently the visualizations produced. These systems vary from service to service and the design of digital dashboards may need to be more ambitious, for example, incorporating solutions to automate data upload to support data timeliness and to ease the workload of those supporting audit participation where necessary. This may be particularly pertinent as health care providers around the world seek to move to paperless systems.

Strengths and Weaknesses

A challenge when evaluating digital technologies is that development is *a continuous cycle* in which users adapt to the technology and adapt the technology over time [36]. Installation of a new iteration of QualDash (Version 2) was disrupted by COVID-19; therefore, it was not possible to capture the impact of this improved version on end use or impact. However, our in-depth data collection across multiple sites and clinical areas enabled us to explore how interactions between the social and technical elements of the system shaped dashboard impact. We were able to explore the reasons behind participants' choice to use or not use QualDash, and the circumstances that influenced those choices, which has generated knowledge that can support the design process for quality dashboards outside the context of this study and strategies to support uptake and adoption.

Conclusions

Web-based, customizable dashboards have the potential to support the use of national audit data in quality improvement, reducing variation in data use across services nationally, and, as COVID-19 has demonstrated, in response to specific events or crises. To optimize the dashboard impact in different settings, co-design would benefit from use of *real* site data, so stakeholders can scrutinize metrics and functions with data that are familiar and meaningful to them. In this way, dashboard development would more closely align with the needs and working practices of key data users within the service. Furthermore, data flow, from collection to upload, needs consideration in the design process, to provide insight into interdependent issues, such as data timeliness and quality, that are likely to impact dashboard use in quality improvement.

Acknowledgments

The authors would like to thank the National Health Service staff who generously gave their time to speak with us and allowed us to observe their practice, without which this research would not be possible. We also thank the School of Health Care, University of Leeds, for their ongoing support in this research. This research was funded by the National Institute for Health Research Health Services and Delivery Research Program (project #16/04/06). The views and opinions expressed are those of the presenter and do not necessarily reflect those of the Health Services and Delivery Research program, National Institute for Health Research, National Health Service or the Department of Health.

Conflicts of Interest

CG is a member of the Myocardial Ischemia National Audit Project Academic and Steering Groups. RF is principal investigator for Pediatric Intensive Care Audit Network. The authors have no other competing interests to declare.

Multimedia Appendix 1
Context + Mechanism = Outcome configurations.
[DOCX File , 14 KB - jmir_v23i11e28854_app1.docx]

References

1. Schoen C, Osborn R, Huynh PT, Doty M, Peugh J, Zapert K. On the front lines of care: primary care doctors' office systems, experiences, and views in seven countries. Health Aff (Millwood) 2006;25(6):w555-w571. [doi: 10.1377/hlthaff.25.w555] [Medline: 17102164]
2. Wunsch H, Angus DC, Harrison DA, Collange O, Fowler R, Hoste EA, et al. Variation in critical care services across North America and Western Europe. Crit Care Med 2008 Oct;36(10):2787-93, e1-9. [doi: 10.1097/CCM.0b013e318186aec8] [Medline: 18766102]
3. Foy R, Skrypak M, Alderson S, Ivers NM, McInerney B, Stoddart J, et al. Revitalising audit and feedback to improve patient care. BMJ 2020 Feb 27;368:m213 [FREE Full text] [doi: 10.1136/bmj.m213] [Medline: 32107249]
4. Allwood D. Engaging clinicians in quality improvement through national clinical audit. Healthcare Quality Improvement Partnership. 2014. URL: https://www.hqip.org.uk/wp-content/uploads/2018/02/engaging-clinicians-in-qi-through-national-clinical-audit.pdf [accessed 2021-11-05]
5. Elshehaly M, Randell R, Brehmer M, McVey L, Alvarado N, Gale CP, et al. QualDash: adaptable generation of visualisation dashboards for healthcare quality improvement. IEEE Trans Vis Comput Graph 2021 Feb;27(2):689-699. [doi: 10.1109/TVCG.2020.3030424] [Medline: 33048727]
6. Taylor A, Neuburger J, Walker K, Cromwell D, Groene O. How is feedback from national clinical audits used? Views from English National Health Service trust audit leads. J Health Serv Res Policy 2016 Apr 24;21(2):91-100. [doi: 10.1177/1355819615612826] [Medline: 26811374]
7. Alvarado N, McVey L, Greenhalgh J, Dowding D, Mamas M, Gale C, et al. Exploring variation in the use of feedback from national clinical audits: a realist investigation. BMC Health Serv Res 2020 Sep 11;20(1):859 [FREE Full text] [doi: 10.1186/s12913-020-05661-0] [Medline: 32917202]
8. McVey L, Alvarado N, Keen J, Greenhalgh J, Mamas M, Gale C, et al. Institutional use of National Clinical Audits by healthcare providers. J Eval Clin Pract 2021 Feb 20;27(1):143-150. [doi: 10.1111/jep.13403] [Medline: 32307857]
9. Hutchinson JW, Alba JW, Eisenstein EM. Heuristics and biases in data-based decision making: effects of experience, training, and graphical data displays. J Mark Res 2010 Aug 01;47(4):627-642 [FREE Full text] [doi: 10.1509/jmkr.47.4.627]
10. van der Meulen M, Logie RH, Freer Y, Sykes C, McIntosh N, Hunter J. When a graph is poorer than 100 words: a comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. Appl Cognit Psychol 2010 Jan;24(1):77-89 [FREE Full text] [doi: 10.1002/acp.1545]
11. Elshehaly M, Alvarado N, McVey L, Randell R, Ruddle R. Requirements specification for a quality dashboard for exploring national clinical audit data. White Rose Research Online. 2019. URL: http://eprints.whiterose.ac.uk/142569/ [accessed 2021-11-05]
12. Dowding D, Merrill J, Russell D. Using feedback intervention theory to guide clinical dashboard design. AMIA Annu Symp Proc 2018 Dec 5;2018:395-403 [FREE Full text] [Medline: 30815079]
13. Wissel B, Van Camp PJ, Kouril M, Weis C, Glauser TA, White PS, et al. An interactive online dashboard for tracking COVID-19 in US counties, cities, and states in real time. J Am Med Inform Assoc 2020 Jul 01;27(7):1121-1125 [FREE Full text] [doi: 10.1093/jamia/ocaa071] [Medline: 32333753]
14. Ivanković D, Barbazza E, Bos V, Brito Fernandes Ó, Jamieson Gilmore K, Jansen T, et al. Features constituting actionable COVID-19 dashboards: descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. J Med Internet Res 2021 Feb 24;23(2):e25682 [FREE Full text] [doi: 10.2196/25682] [Medline: 33577467]

15. NICOR. Case Ascertainment for Trusts in England. London: National Institute for Cardiovascular Outcomes Research (NICOR); 2021.
16. Wilkinson C, Weston C, Timmis A, Quinn T, Keys A, Gale CP. The Myocardial Ischaemia National Audit Project (MINAP). Eur Heart J Qual Care Clin Outcomes 2020 Jan 01;6(1):19-22. [doi: 10.1093/ehjqcco/qcz052] [Medline: 31511861]
17. PICANet 2015 Annual Report. Healthcare Quality Improvement Partnership. 2015. URL: https://www.picanet.org.uk/wp-content/uploads/sites/25/2018/05/PICANet_2015_Annual_Report_Summary.pdf [accessed 2021-10-25]
18. Elshehaly M, Randell R, Brehmer M, McVey L, Alvarado N, Gale CP, et al. QualDash: adaptable generation of visualisation dashboards for healthcare quality improvement. IEEE Trans Vis Comput Graph 2021 Feb;27(2):689-699. [doi: 10.1109/TVCG.2020.3030424] [Medline: 33048727]
19. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
20. Lawler EK, Hedge A, Pavlovic-Veselinovic S. Cognitive ergonomics, socio-technical systems, and the impact of healthcare information technologies. Int J Ind Ergon 2011 Jul;41(4):336-344. [doi: 10.1016/j.ergon.2011.02.006]
21. Sittig D, Singh H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. In: Cognitive Informatics for Biomedicine. Cham: Springer; 2015:59-80.
22. Fox WM. Sociotechnical system principles and guidelines: past and present. J Appl Behav Sci 2016 Jul 26;31(1):91-105. [doi: 10.1177/0021886395311009]
23. Pawson R, Tilley N. An introduction to scientific realist evaluation. In: Evaluation for the 21st Century: A Handbook. Thousand Oaks: SAGE Publications; 1997.
24. Randell R, Greenhalgh J, Dowding D. Using realist reviews to understand how health IT works, for whom, and in what circumstances. J Am Med Inform Assoc 2015 Apr;22(e1):e216-e217 [FREE Full text] [doi: 10.1093/jamia/ocu006] [Medline: 25755123]
25. Goldsmith L. Using framework analysis in applied qualitative research. Qual Rep 2021 Jun 20;26(6):2061-2076 [FREE Full text] [doi: 10.46743/2160-3715/2021.5011]
26. Parkinson S, Eatough V, Holmes J, Stapley E, Midgley N. Framework analysis: a worked example of a study exploring young people's experiences of depression. Qual Res Psychol 2015 Nov 16;13(2):109-129. [doi: 10.1080/14780887.2015.1119228]
27. Srivastava A, Thomson SB. Framework analysis: a qualitative methodology for applied policy research. J Administration Govern 2009;4(2):72-79 [FREE Full text]
28. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess R, editors. Analyzing Qualitative Data. London: Routledge; 1994:173-194.
29. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
30. Petracca F, Ciani O, Cucciniello M, Tarricone R. Harnessing digital health technologies during and after the COVID-19 pandemic: context matters. J Med Internet Res 2020 Dec 30;22(12):e21815 [FREE Full text] [doi: 10.2196/21815] [Medline: 33351777]
31. Lian W, Wen L, Zhou Q, Zhu W, Duan W, Xiao X, et al. Digital health technologies respond to the COVID-19 pandemic in a tertiary hospital in China: development and usability study. J Med Internet Res 2020 Nov 24;22(11):e24505 [FREE Full text] [doi: 10.2196/24505] [Medline: 33141679]
32. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015 Feb;84(2):87-100. [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
33. Janssen A, Donnelly C, Kay J, Thiem P, Saavedra A, Pathmanathan N, et al. Developing an intranet-based lymphedema dashboard for breast cancer multidisciplinary teams: design research study. J Med Internet Res 2020 Apr 21;22(4):e13188 [FREE Full text] [doi: 10.2196/13188] [Medline: 32314968]
34. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22 [FREE Full text] [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
35. Richter Lagha R, Burningham Z, Sauer BC, Leng J, Peters C, Huynh T, et al. Usability testing a potentially inappropriate medication dashboard: a core component of the dashboard development process. Appl Clin Inform 2020 Aug 12;11(4):528-534 [FREE Full text] [doi: 10.1055/s-0040-1714693] [Medline: 32785904]
36. Carayon P. Human factors in health(care) informatics: toward continuous sociotechnical system design. Stud Health Technol Inform 2019 Aug 09;265:22-27 [FREE Full text] [doi: 10.3233/SHTI190131] [Medline: 31431572]

Abbreviations

CMOc: Context + Mechanism = Outcome configuration
IT: information technology
MINAP: Myocardial Ischemia National Audit Project
NHS: National Health Service

PICANet: Pediatric Intensive Care Audit Network
PICU: pediatric intensive care unit
TAM: technology acceptance model

Edited by R Kukafka, G Eysenbach; submitted 14.04.21; peer-reviewed by J Chishtie, I Wilson, J Oldenburg; comments to author 28.06.21; revised version received 15.07.21; accepted 05.10.21; published 23.11.21.

Please cite as:
Alvarado N, McVey L, Elshehaly M, Greenhalgh J, Dowding D, Ruddle R, Gale CP, Mamas M, Doherty P, West R, Feltbower R, Randell R
Analysis of a Web-Based Dashboard to Support the Use of National Audit Data in Quality Improvement: Realist Evaluation
J Med Internet Res 2021;23(11):e28854
URL: https://www.jmir.org/2021/11/e28854
doi:10.2196/28854
PMID:34817384

Original Paper

Developing the Total Health Profile, a Generalizable Unified Set of Multimorbidity Risk Scores Derived From Machine Learning for Broad Patient Populations: Retrospective Cohort Study

Abhishaike Mahajan[1], BS; Andrew Deonarine[2], MS, MD, PhD; Axel Bernal[1], PhD; Genevieve Lyons[2], MS; Beau Norgeot[1], PhD

[1]Anthem Inc, Palo Alto, CA, United States

[2]XY Health, Cambridge, MA, United States

Corresponding Author:
Beau Norgeot, PhD
Anthem Inc
661 Bryant St
Palo Alto, CA, 94301
United States
Phone: 1 650 272 7314
Email: beaunorgeot@gmail.com

Abstract

Background: Multimorbidity clinical risk scores allow clinicians to quickly assess their patients' health for decision making, often for recommendation to care management programs. However, these scores are limited by several issues: existing multimorbidity scores (1) are generally limited to one data group (eg, diagnoses, labs) and may be missing vital information, (2) are usually limited to specific demographic groups (eg, age), and (3) do not formally provide any granularity in the form of more nuanced multimorbidity risk scores to direct clinician attention.

Objective: Using diagnosis, lab, prescription, procedure, and demographic data from electronic health records (EHRs), we developed a physiologically diverse and generalizable set of multimorbidity risk scores.

Methods: Using EHR data from a nationwide cohort of patients, we developed the total health profile, a set of six integrated risk scores reflecting five distinct organ systems and overall health. We selected the occurrence of an inpatient hospital visitation over a 2-year follow-up window, attributable to specific organ systems, as our risk endpoint. Using a physician-curated set of features, we trained six machine learning models on 794,294 patients to predict the calibrated probability of the aforementioned endpoint, producing risk scores for heart, lung, neuro, kidney, and digestive functions and a sixth score for combined risk. We evaluated the scores using a held-out test cohort of 198,574 patients.

Results: Study patients closely matched national census averages, with a median age of 41 years, a median income of $66,829, and racial averages by zip code of 73.8% White, 5.9% Asian, and 11.9% African American. All models were well calibrated and demonstrated strong performance with areas under the receiver operating curve (AUROCs) of 0.83 for the total health score (THS), 0.89 for heart, 0.86 for lung, 0.84 for neuro, 0.90 for kidney, and 0.83 for digestive functions. There was consistent performance of this scoring system across sexes, diverse patient ages, and zip code income levels. Each model learned to generate predictions by focusing on appropriate clinically relevant patient features, such as heart-related hospitalizations and chronic hypertension diagnosis for the heart model. The THS outperformed the other commonly used multimorbidity scoring systems, specifically the Charlson Comorbidity Index (CCI) and the Elixhauser Comorbidity Index (ECI) overall (AUROCs: THS=0.823, CCI=0.735, ECI=0.649) as well as for every age, sex, and income bracket. Performance improvements were most pronounced for middle-aged and lower-income subgroups. Ablation tests using only diagnosis, prescription, social determinants of health, and lab feature groups, while retaining procedure-related features, showed that the combination of feature groups has the best predictive performance, though only marginally better than the diagnosis-only model on at-risk groups.

Conclusions: Massive retrospective EHR data sets have made it possible to use machine learning to build practical multimorbidity risk scores that are highly predictive, personalizable, intuitive to explain, and generalizable across diverse patient populations.

(J Med Internet Res 2021;23(11):e32900) doi:10.2196/32900

KEYWORDS

multimorbidity; clinical risk score; outcome research; machine learning; electronic health record; clinical informatics; morbidity; risk; outcome; population data; diagnostic; demographic; decision making; cohort; prediction

Introduction

Multimorbidity risk scores, which factor in the presence of several chronic conditions, can provide insights into morbidity and mortality [1,2]. In general, the number of co-occurring medical conditions is associated with increased adverse medical outcomes [3-5] and increased use of medical services [6]. This is particularly true for older individuals since the number of co-occurring medical conditions will increase with age [7]. Various approaches to quantifying multimorbidity have been used, including simply counting the number of conditions [8], while more complex tools have also been developed, such as the Charlson Comorbidity Index (CCI) [9] and the Elixhauser Comorbidity Index (ECI) [10]. However, these scores were developed decades ago and are limited to diagnostic information or certain populations, which may limit their utility. A systematic review [11] of multimorbidity scores examined 35 major multimorbidity scores, which could be subclassified by the information they used (eg, prescription data, diagnostic data, self-reported quality of life) and the outcomes they recorded (all-cause mortality, emergency room admissions, and hospital admissions). Patients with multimorbidities are cared for in general practice and by specialists [8,12,13] who use disease-specific risk scores and guidelines. Condition-specific risk scores, such as the Framingham Risk Score [14] for coronary heart disease, can help identify specific interventions to benefit patients and provide actionable information to guide tests and medications. One potential issue with the use of these tools is that sometimes they do not address the overall health care priorities of the patient due to their narrow focus [8,15,16].

Multimorbidity scores tend to only use one type of clinical data, such as diagnoses, prescriptions, or procedures, and rarely integrate them. As a result, they may be missing vital information and relationships in patient information. Although newer methods, such as probabilistic phenotyping [17], may alleviate these concerns, while remaining scalable, these methods are still highly experimental, with a wide variety of methods and little consensus on which ones are most trustable for real-world settings [18]. Using multiple data sets, feature types, and methodological explorations could provide a more comprehensive and robust estimate of multimorbidity risk. Currently, no multimorbidity scores exist that produce granular and overall risk profiles irrespective of age and sex; are derived from a large, representative population of patients; and integrate multiple clinical data sets, including diagnoses, prescriptions, lab results, and procedures using machine learning (ML; building upon previous ML-based strategies and recommendations for multimorbidity analysis by Hassaine et al [19]). Such scores could help health care providers engage in patient-centered care and prescribing, reduce polypharmacy, and guide deprescribing when used together with traditional risk scores and guidelines.

To address this need, we sought to create the total health profile (THP), a set of ML-derived measures of an individual's comprehensive clinical risk. The THP presents clinical risk in five separate models (referred to as "component scores"), producing granular, multimorbid risk scores specific to cardiovascular ("heart score"), respiratory ("lung score"), neuropsychiatric ("neuro score"), renal ("kidney score"), and gastrointestinal ("digestive score") conditions. These organ systems reflect those involved in the top 10 sources of mortality in the United States [20] and serve to complement existing disease-specific risk scores. We also included, as a sixth score, the total health score (THS), a single view of a patient's overall health across all five of the aforementioned organ systems, which can be compared to existing pure multimorbidity risk scores. Each of these six scores was independently modeled using electronic health record (EHR) data consisting of demographic information, diagnosis codes, lab results, prescriptions, and medical procedural data and required, otherwise, no patient behavior or familial history data. For the unified risk endpoint of all six of the scoring models, we used inpatient (IP) hospital visits. As such, each score's estimate of clinical risk represents the likelihood of an IP hospital visit over the next 24 months, attributable to the score's clinical category (eg, lung, heart). After training, testing, and calibrating the THS and the five organ system component scores, we analyzed the metrics and generalizability of each score across populations. We also conducted ablation tests of several feature groups to assess their importance in the final set of models. Finally, we discussed the clinical applicability of the THP, limitations of the study, and future work.

Methods

Study Design and Patient Inclusion Criteria

This retrospective cohort study used lab measurements and an administrative claims database of 52 million patients provided by a US health care insurance company. Patients were enrolled in a mixture of commercial, Medicare, Medicaid, and exchange plans. Our study design involved training on retrospective data from a certain time window and assessing performance via a follow-up time window. The retrospective observation period, or the time period in which model features were collected, was defined as January 1, 2016, through December 1, 2017, and the follow-up period, or the time period in which the model labels were collected, as January 1, 2018, through December 31, 2019.

All patient data were de-identified. Patients selected for inclusion had at least one medical claim in each year of the data collection and follow-up periods and had a known sex, birthdate, and zip code. These inclusion criteria resulted in 14 million patients, from which 1 million patients were randomly selected for analysis using PySpark, resulting in 992,868 patients due to the approximation methods used by PySpark. This patient sample was split into training (n=794,294) and testing (n=198,574) groups corresponding to an 80:20 ratio. Diagnosis codes (*International Classification of Diseases, Tenth Revision* [ICD-10]), medical procedure codes (Current Procedural

Terminology [CPT]), lab data, demographics (social determinants of health [SDoH], patient gender/age), and prescription data (defined by General Product Identifier [GPI] codes) were used from patients who met the selection criteria. Our study, in total, used 88 ICD-10 codes and 30 chronic conditions (derived from ICD-10 codes specified by the ECI), 16 lab types, 764 GPI codes representing 4 GPI prefixes, 14 CPT codes, and 17 demographic markers.

Data Processing

Using the data compiled for the 992,868 patients, we extracted a set of features corresponding to chronic diagnoses, acute diagnoses, IP hospital visits, prescriptions, sociodemographic information, and lab results/physical exam measurements for feature extraction and modeling. A description of all features gathered during the data collection period follows next.

Demographic information was extracted from the United States Census American Community Survey for 2017 at the zip code level. This information included population, household count, and race and ethnicity percentages for that zip code (eg, African American, non-Hispanic White, Hispanic, Asian, Native American), sex percentages per zip code, and economic indicators, including the mean and median income. Demographic data also included the age and sex of the patients. Chronic disease diagnoses were counted as the presence of a chronic disease, while acute diagnoses were counted as the number of those diagnoses in the study period, summed over the component. For instance, 3 atrial fibrillation codes and 2 acute heart failure codes during the 2-year data collection period would have resulted in the number of acute heart diagnoses being 5. Medical procedure features were counted as the count of IP CPT codes that occurred during the data collection period, with otherwise identical score-specific inclusion criteria to the IP hospital labels (discussed in the Model Outcome Labels section). Four groups of prescriptions were included, assigned using the first two digits of the GPI code and indicated by binary presence: antihypertensives, hypoglycemics, lipid-lowering medications, and antithrombotic agents. In all, 16 labs or vitals were included in the study, each one being a numerical feature. If there were multiple results of the same lab data/vitals collected during the data collection period, only the most recent measurement was included. In total, our feature set and labels used the following set of clinical features: diagnoses (88 ICD-10 codes and 30 chronic conditions), labs (16 types), prescriptions (764 GPI codes, representing 4 GPI prefixes), procedures (14 CPT codes), and demographics (15 SDoH and 2 individual patient characteristics).

Except for demographic features, lab values, and vital signs, all input features were filtered on a model-by-model basis to include only score-relevant data (ie, the heart score would be modeled using only physician-curated features related to cardiovascular health). For IP hospital visit features collected during the data collection period, only score-specific IP visit counts were included (ie, the heart score would have as input the number of heart-related IP visitations during the data collection period, not the lung-related, and so on). The set of input features used over all component score models were used as input for the THS model, with an exception for chronic diagnosis features.

Model Outcome Labels

All component score labels were a binary indicator referring to whether a patient had an IP visit within the follow-up period, given that they also had acute or chronic diagnoses within 12 months prior to the IP visit and within 7 days after the IP visit, establishing both a history of that condition and that the IP visit was (likely) related to that condition. These diagnoses would be specific to each component, given by the corresponding ECI comorbidities and ICD-10 codes. For example, a possible positive label for the lung scoring model could be an IP hospital stay CPT code on June 2, 2019, a diagnosis code corresponding to pneumonia 3 months prior to it, and a diagnosis code corresponding to chronic pulmonary disease 2 days after it. The THS label is simply the combination of all the component score labels; if a patient has any positive component score label, the THS label would be positive as well.

Modeling Procedures and Baselines

All scores were calculated using a gradient-boosted tree classifier, with default hyperparameters, using the Scikit-Learn Python 3.6 package (version 0.24.1). Using demographics, diagnoses, lab values, procedures, and prescription data as input and IP visits as binary labels, separate models were trained for each score and subsequently calibrated using an isotonic regression with 3-fold cross-validation over the training set. Discriminative results from the models were obtained using the optimal threshold point of the training set (given by the threshold that yielded the smallest difference between the true-positive rate and the false-positive-rate) and applied to the testing set. All missing values were mean-imputed, and all input features for each model were mean-normalized using the training data.

We had multiple baselines: a logistic regression model with default hyperparameters using the *statsmodel* package (version 0.12.0) with otherwise identical feature sets, and a comparison of the performance of the THS to commonly used scores of a similar nature, specifically the CCI and the ECI, in predicting the hospitalization endpoint. We also conducted multiple ablation tests on the feature groups: a set of gradient-boosted tree classifiers, all with procedural data, but having only one set of either lab, SDoH, prescription, or diagnosis information. For the baseline gradient-boosted comparison with the combined feature model for patient subgroups, CI calculations were generated using 100 bootstrap iterations of 10% of the given demographic. The patient subgroups analyzed were patients with two or more of any comorbidity and one or more prescriptions of hypertensive, hyperglycemic, lipid-lowering, or antithrombotic medications.

Radial Plots

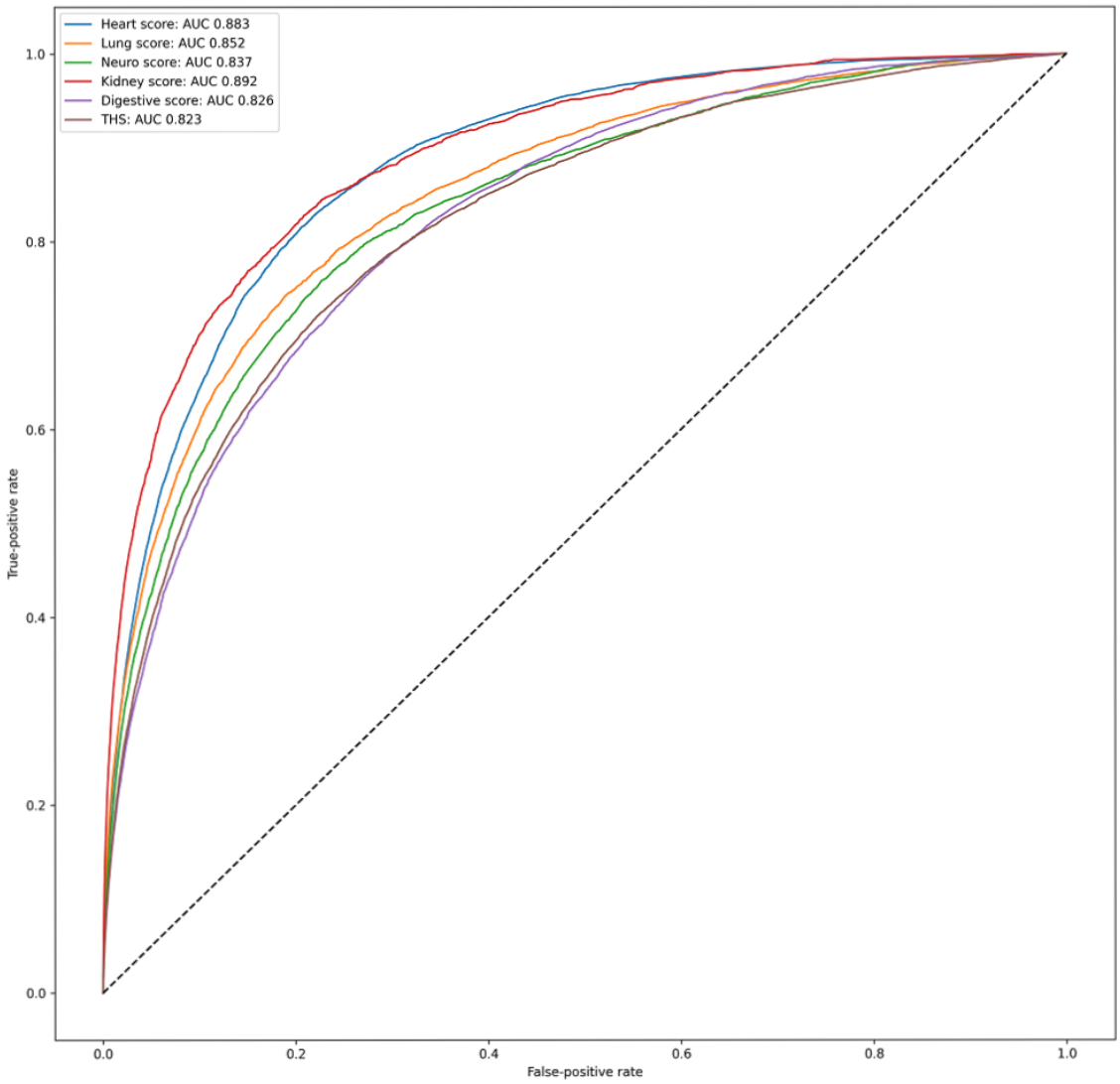
Radial plots were generated using three patients who were closest to each of the centroids of a fitted, randomly initialized K-means model, with a K value of 3. The K-means algorithm used the Scikit-Learn Python 3.6 package (version 0.24.1), and the plots themselves were generated using Plotly.

Model Discrimination and Generalization/Sensitivity Analyses

Models were assessed on three levels: discriminative performance, calibration, and generalizability in performance

across different demographics. To assess the discriminative performance of each model in the THP, we calculated the area under the receiver operating curve (AUROC), sensitivity, and specificity using Scikit-Learn for the testing set of 198,574 patients. We also plotted the AUROC for all scores on the testing set (Figure 1). All CIs for the discriminative metrics were generated using 500 bootstrap samples of 20,000 from the testing data set. We selected the AUROC as our primary metric because it represents a comprehensive measure of the true-positive-rate and false-positive-rate trade-off without needing an optimal threshold point. Since our outcomes exhibited strong class imbalance, which may have led to overly optimistic AUROC values, we used sensitivity and specificity as secondary model measures. To assess calibration performance, we created calibration plots using Scikit-Learn, graphing predicted probability versus positive label percentage across 10 uniform probability bins. We assessed calibration, as measured by calibration plots, as the primary measurement of clinical utility as it gives a clear idea of how these scores can be used to identify sick patients, avoid alarm fatigue, and be interpreted as a probabilistic likelihood. To assess the generalization performance of each model, we studied how the performance and scores of the models vary across age, zip code income, and sex categories. We plotted how the AUROC varies among age groups (decade age groupings), median income groups (low, medium and high), and binary gender (male or female). Additionally, we computed statistical significance Z-tests for AUROC pairwise differences between all groups within each category. Due to the lower sample sizes of the groups, CI calculations were generated using 100 bootstrap samples of 10% of the given demographic of the testing group. The THS and component scores were then analyzed by plotting the distribution of scores as a function of age and disease burden (measured by the presence of pre-existing comorbidities during the observation period). Specifically, we looked at the distributions of the THS and the component scores among various age groups for patients with zero comorbidities found during the data collection period and patients with at least one ECI comorbidity related to the given component found during the data collection period.

Figure 1. ROCs for all scores in the THP. AUC: area under the curve; ROC: receiver operating curve; THP: total health profile; THS: total health score.



Physician-Guided Feature Selection and Curation

To select features to incorporate into the THS, a physician-guided curation method was incorporated, which involved selecting conditions, reviewing clinical practice guidelines for important conditions, and identifying clinical measures, tests, and pharmacological interventions in those guidelines. No statistical feature selection techniques were used, as those offer improved accuracy mainly in cases with relatively small training data sets or models that are sensitive to unsuspected feature correlations [21].

An overview of the manually guided feature selection process is described below:

1. Selection of disease categories/subscores: The top causes of death across the United States were reviewed from the 2018 mortality statistics from the National Center for Health Statistics, and five main categories that the causes of death could be classified into were identified: cardiac (heart), respiratory (lung), neuropsychiatric (neuro), gastrointestinal (digestive), and renal (kidney).
2. We then obtained the leading causes of IP conditions for 2011-2013 using the Agency for Health care Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) database. We cross-referenced the top 30 codes for each year with the 5 categories we developed in step 1. We did not include codes corresponding to obstetrical conditions, complications related to birth and delivery, multisystem malignancy, and musculoskeletal conditions.
3. Using the AHRQ HCUP codes, we extracted the ICD-10 codes and selected additional medically related conditions (eg, selecting the ICD-10 codes for ischemic stroke codes, in addition to ICD-10 codes for hemorrhagic stroke).
4. To obtain corresponding prescription, lab, and procedure codes, we then reviewed clinical practice guidelines for the identified conditions (eg, stroke, chronic obstructive pulmonary disease) from the United States Preventive Services Task Force [18,22]. These guidelines were then reviewed and lab data and procedures corresponding to diagnosis and management were extracted by a physician. These were then manually mapped to the corresponding drug (GPI codes), procedure (CPT codes), and lab (Logical Observation Identifiers Names and Codes) codes.

MI-CLAIM Checklist

This work meets the Minimum Information about Clinical Artificial Intelligence Modeling (MI-CLAIM) requirements for sharing design, data/optimization, model performance, model examination, and level of reproducibility [23].

Role of the Funding Source

The funding source collected the raw data.

Ethics Approval

The analysis presented here is not to be characterized as human subject research. We are presenting the results of an analysis conducted for a health plan's health care operations in accordance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Only aggregated results of the business analysis are provided, and no individually identifiable information (protected health information or otherwise) was used in the development of this presentation.

Results

Patient Cohort

In all, 992,868 patients matched the inclusion criteria (Table 1). The majority were female (n=560,165, 56.4%), with 432,703 (43.6%) males, concordant with the 2019 census results [20]; the median age (41 years) was slightly higher than the national average; and the number of comorbidities tended to increase with age, consistent with previous findings [7]. The mean patient age was 39.07 years (95% CI 39.02-39.12), the mean number of comorbidities was 1.71 (95% CI 1.71-1.72), and the percentage of patients with IP visits was 1.65% (females 1.7%, males 1.7%). IP visits were also positively correlated with age for each organ system (Spearman correlation=0.314), which is concordant with previous studies [21]. The sample-weighted summary of zip-code-level demographics had a median income of $66,829, and the racial averages by zip code were 73.8% White, 5.9% Asian, and 11.9% African American, closely matching census averages (Table 1).

Table 1. Demographic profile of patients included in the analysis cohort (N=992,868).

Characteristics	Overall	Females	Males
Age range, n (%)			
0-10 years	150,685 (15.2)	73,224 (48.6)	77,461 (51.4)
10-20 years	140,684 (14.2)	73,657 (52.4)	67,027 (47.6)
20-30 years	80,136 (8.1)	55,013 (68.6)	25,123 (31.4)
30-40 years	102,397 (10.3)	65,797 (64.3)	36,600 (35.7)
40-50 years	126,923 (12.8)	75,151 (59.2)	51,772 (40.8)
50-60 years	163,675 (16.5)	91,305 (55.8)	72,370 (44.2)
60-70 years	122,243 (12.3)	66,270 (54.2)	55,973 (45.8)
70-80 years	72,560 (7.3)	39,633 (54.6)	32,927 (45.4)
80-90 years	33,565 (3.4)	20,115 (59.9)	13,450 (40.1)
Number of comorbidities, n (%)			
0	362,469 (36.5)	202,023 (55.7)	160,446 (44.3)
1	241,166 (24.2)	134,245 (55.7)	106,921 (44.3)
2	140,558 (14.1)	81,174 (57.8)	59,384 (42.2)
3	89,258 (8.9)	51,827 (58.1)	37,431 (41.9)
4+	159,417 (16.0)	90,896 (57.0)	68,521 (43.0)
Zip code demographics (Census 2019)			
% White (mean %)	73.8 (76)	73.4% (N/A[a])	74.3% (N/A)
% Black (mean %)	11.9 (13)	12.3% (N/A)	11.4% (N/A)
% Asian (mean %)	5.9 (5)	5.8% (N/A)	6.0% (N/A)
Median income (mean)	$66,829 ($62,843)	$66,431 (N/A)	$67,343 (N/A)

[a]N/A: not available.

Overall Model Performance

All models outperformed the logistic regression baseline and were well specified with AUROCs of 0.83 for the THS, 0.89 for heart, 0.86 for lung, 0.84 for neuro, 0.90 for kidney, and 0.83 for digestive functions (Figure 1). All six models were well calibrated. Additional metrics (sensitivity and specificity) can be found in Table 2.

One benefit of the THP is that it is personalized to the patient to allow for nuanced interpretation based on the affected organ system, in addition to robust predictive performance. Figure 2 demonstrates an illustrative radial plot example of three patients who were around the same age (50-60 years old) and had the same rough THS (>0.8). Unlike grouped scoring systems, the THP enables the clinician to understand the personalized drivers for that score, thereby enabling clinical decisions that are specific to the individual patient. The score of patient 1 was driven primarily by neuro and heart issues, while the score of patient 2 was affected by kidney, neuro, and heart diseases, and the score of patient 3 was mostly affected by heart, lung, digestive, and kidney maladies.

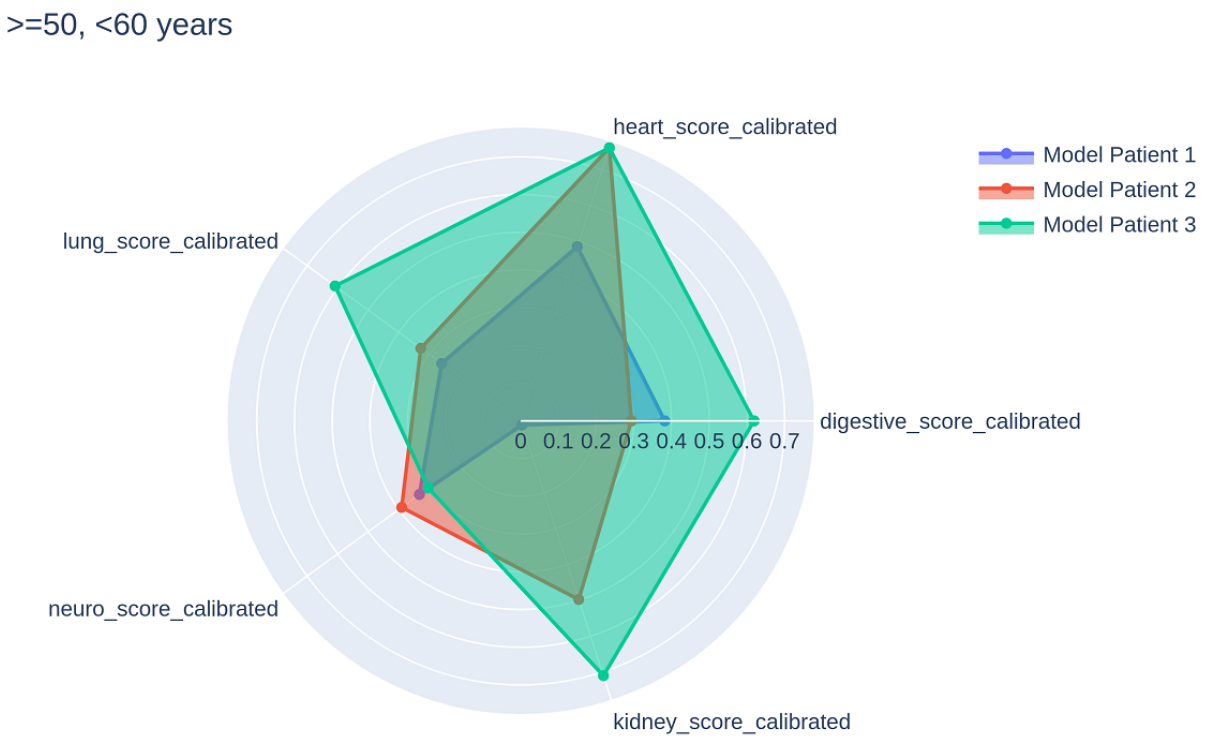
Table 2. Gradient-boosted tree AUROC[a], sensitivity, and specificity for each score in the testing set (n=198,574).

Score type	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Heart	0.883 (0.876-0.893)	0.82 (0.796-0.845)	0.788 (0.783-0.793)
Lung	0.853 (0.837-0.867)	0.75 (0.713-0.784)	0.802 (0.796-0.808)
Neuro	0.837 (0.821-0.855)	0.756 (0.722-0.793)	0.774 (0.768-0.78)
Kidney	0.892 (0.873-0.908)	0.784 (0.738-0.825)	0.83 (0.824-0.835)
Digestive	0.827 (0.81-0.847)	0.733 (0.698-0.767)	0.756 (0.75-0.762)
THS[b]	0.823 (0.811-0.834)	0.721 (0.701-0.744)	0.777 (0.771-0.783)

[a]AUROC: area under the receiver operating curve.

[b]THS: total health score.

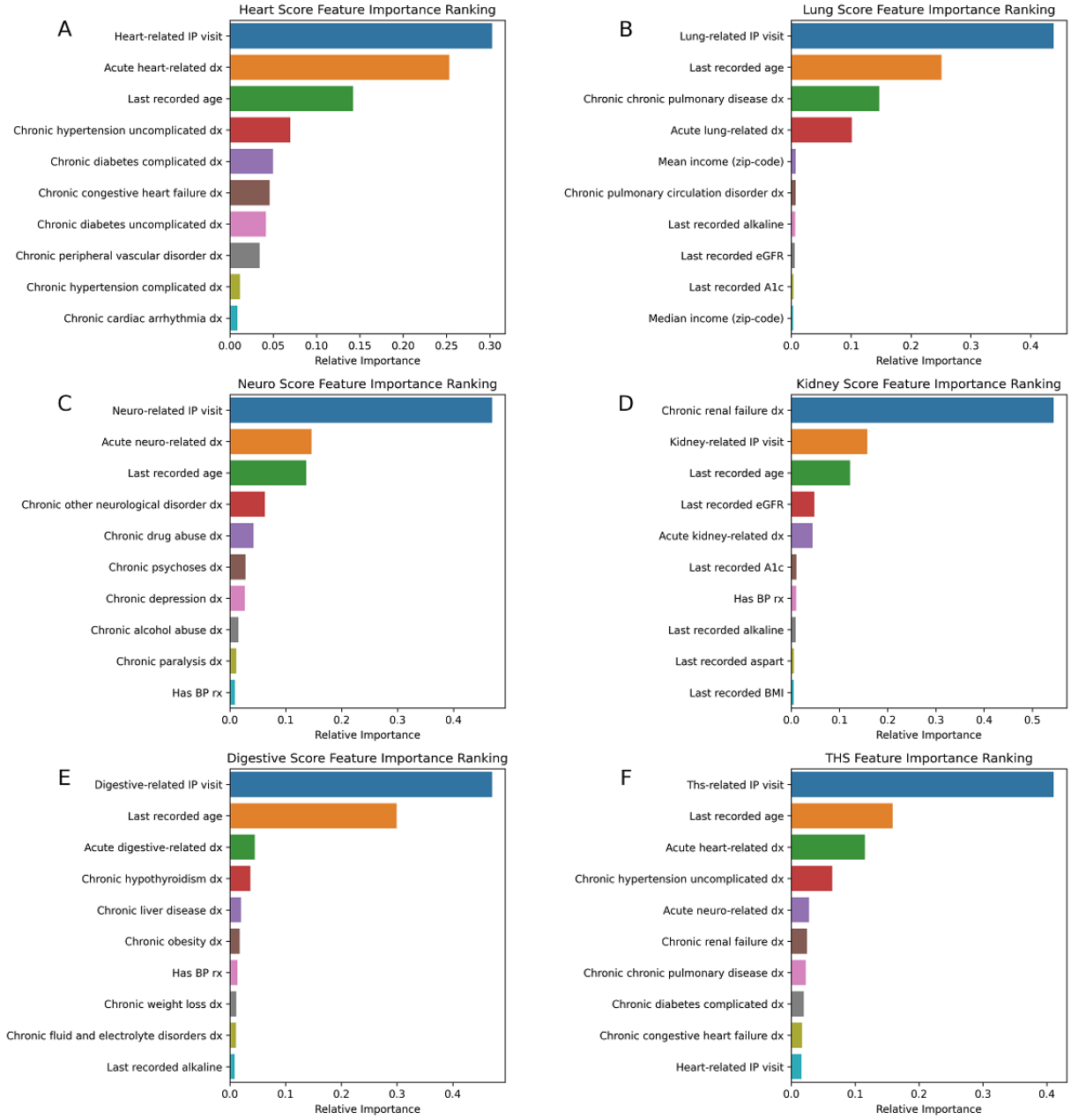
Figure 2. Three patients, all 50-60 years of age and all with approximately equal THSs. THS: total health score.



Important Model Features

We obtained the most important features of the THS and of each component model. For the THS, the biggest drivers of the model were the use of prior IP hospital visits (0.41) and age (0.15), followed by an acute heart-related diagnosis (0.11), uncomplicated hypertension (0.06), and acute neurological conditions (0.02); see Figure 3. The key features for each component model were directly relevant elements. For example, the most important features of the kidney model included a diagnosis of renal failure, as well as age, last recorded measurement of estimated glomerular filtration rate (eGFR) and hemoglobin A_{1c} (HbA_{1c}), and any acute kidney-related diagnosis. Generally, age and a prior history of hospitalization for issues relating to the organ system in question were the most important features across most component models.

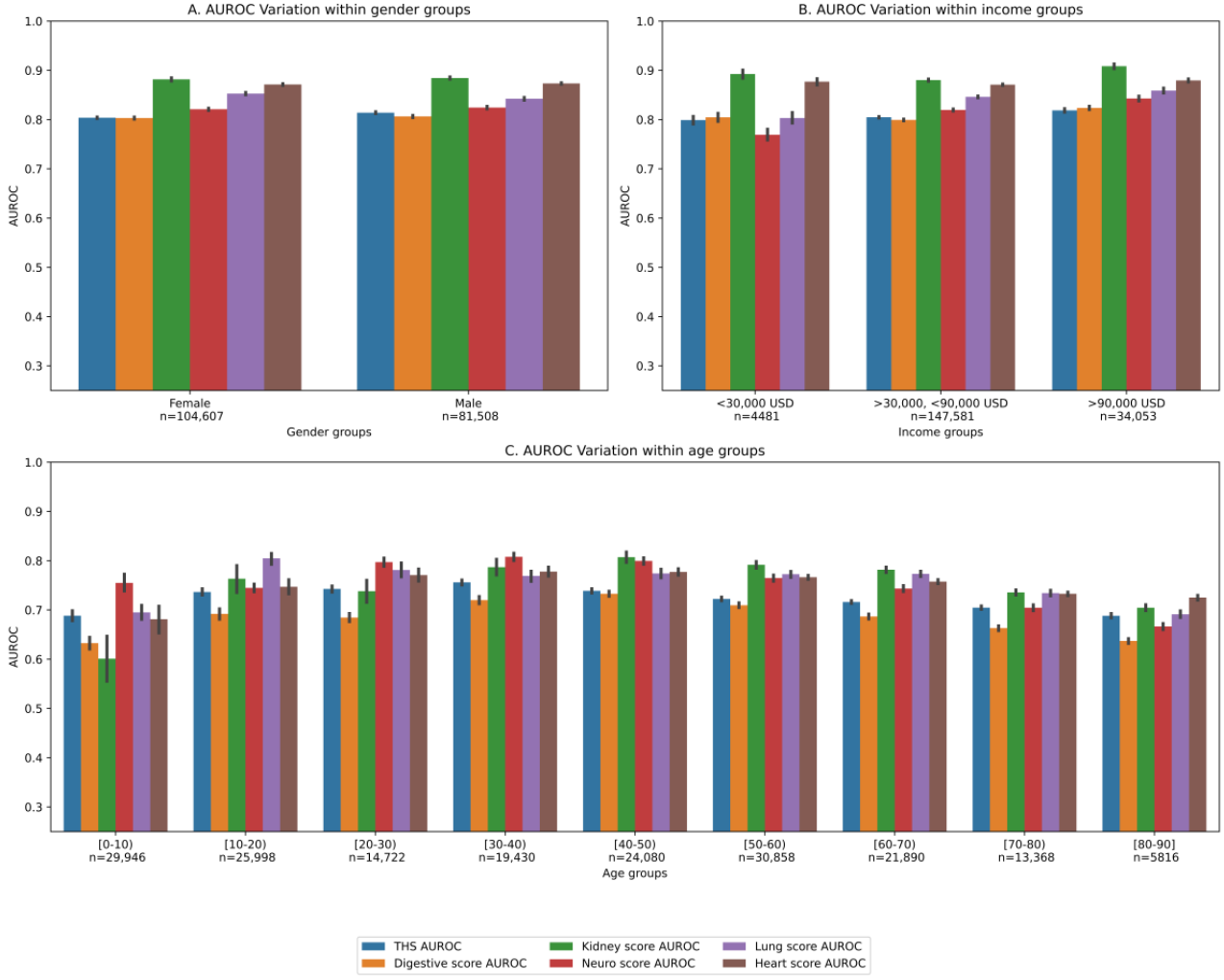
Figure 3. Feature importance plots across all scoring models, generated using Gini impurity reduction. BMI: body mass index; BP, blood pressure; dx: diagnosis; eGFR: estimated glomerular filtration rate; HbA_{1c}: hemoglobin A_{1c}; IP: inpatient; THS: total health score.



Generalizability Across Subgroups

We examined differences in performance for the THS as well as each component score for population subgroups based on sex, median neighborhood income, and age (Figures 4A, 4B, and 4C, respectively). Statistical AUROC comparisons [24] with Bonferroni corrections revealed that there were no significant differences in model performance on the basis of sex or income. There were occasional statistical differences in performance on the basis of age, primarily related to the 80-90-year-old population, which had significantly fewer data points than any of the other age groups (see Table 1).

Figure 4. Generalizability. AUROC performance across population subgroups. AUROC: area under the receiver operating curve; THS: total health score.



Ablation Tests

We conducted ablation tests of several feature groups to assess the need for them in the final set of models. All ablation test models used procedural data as input, as we assumed access to this information is a given due to the outcome prediction being procedural as well. The combined feature set outperformed the lab, SDoH, and prescription-only models. The combined feature set also outperformed the diagnosis-only model, though not significantly. This statistically nonsignificant outperformance was similarly observed in multiple patient subgroups focused on at-risk patients.

Performance Comparisons to Charlson and Elixhauser Risk Scores

In addition to a model baseline (logistic regression), we also compared the performance of the THS to commonly used scores of a similar nature, specifically the CCI and the ECI. Both the CCI and the ECI generate a risk score based on different weights associated with certain diseases based on ICD-10 codes. The THS (AUROC=0.82) outperformed both the CCI (AUROC=0.74) and the ECI (AUROC=0.65); see Table 3. We further compared performance across subgroups. Of the baselines, the CCI is a consistently better predictor than the ECI. Across subgroups, the THS also consistently outperformed the score baselines for every age, sex, and income bracket. The improvement was perhaps most pronounced for ages between 20 and 50, as well as for individuals living in lower-income neighborhoods.

Table 3. AUROC[a] performance by sociodemographic strata and score in predicting IP[b] visitations.

Strata	ECI[c] score	CCI[d] score	THS[e] feature score
All	0.649	0.735	0.823
Gender[f]			
0	0.637	0.733	0.822
1	0.664	0.736	0.824
Age bracket			
0-10 years	0.621	0.622	0.696
20-30 years	0.484	0.573	0.725
20-30 years	0.460	0.621	0.786
30-40 years	0.509	0.635	0.768
40-50 years	0.549	0.675	0.769
50-60 years	0.602	0.693	0.754
60-70 years	0.637	0.694	0.742
70-80 years	0.645	0.678	0.721
80+ years	0.652	0.657	0.711
Income, median			
$0-$30,000	0.640	0.715	0.820
$30,000-$90,000	0.647	0.734	0.823
$90,000+	0.665	0.738	0.833

[a]AUROC: area under the receiver operating curve.

[b]IP: inpatient.

[c]CI: Elixhauser Comorbidity Index.

[d]CCI: Charlson Comorbidity Index.

[e]THS: total health score.

[f]0: male; 1: female.

Discussion

Principal Findings

There is a continued need for an updated clinical score that profiles patients based on multimorbidities that are equitable across populations and nuanced enough to facilitate precision medicine. To facilitate clinical decision making across patient populations, we created an automated, generalizable, integrated, multimorbidity risk profile across several clinical domains. The THP is composed of cardiovascular, respiratory, neuropsychiatric, renal, and gastrointestinal clinical risk subprofiles, as well as a sixth score, the THS, representing the overall combinatorial risk of the five organ-specific scores. We followed ML best practices to train six integrated models on large-scale EHR data with the forecasted probability of a risk endpoint, organ-specific IP hospital visits, over a 2-year window as the target. We chose IP hospital stays as our risk endpoint because reductions in overall health, whether due to multiple health conditions [6,25] or aging [26], are associated with increased hospital visits [27,28]. The primary contribution of this work goes beyond the models themselves by matching clinical knowledge to data that are available at scale, across a diverse cohort of patients. In our experiments, we found that the profile demonstrates high performance in terms of the AUROC on the aggregate held-out testing set. Importantly, there was consistent performance of this scoring system across sexes, diverse patient ages, and income levels. The THS model and each of the component models learned to generate predictions by focusing on appropriate, clinically relevant patient features. The THP is personalized based on individual organ system risk drivers, and visualizations, such as radar plots, can be used to facilitate explainability and encourage confidence of clinical decision making, providing meaningful feature importance. The THS outperformed relevant baselines, specifically the other commonly used multimorbidity scoring systems CCI and ECI, for every age, sex, and income bracket. Finally, we conducted multiple ablation tests, while retaining procedure features, to determine the relative contribution of feature groups to the THP. In this experiment, we found that while the combined feature set predictive performance outperformed the prescription, lab, and SDoH ablations, it was largely similar to the diagnosis ablation. However, we hypothesized that we would find larger differences in performance among at-risk populations and found, in patients with multiple comorbidities and on certain prescriptions, minor but consistent increases in predictive power using the combined feature set versus the diagnosis feature set, implying that risk

prediction is improved on more complex patients, given more complex data. As more features, including more labs, diagnosis, and prescriptions, are added to the THP, future work will more closely examine which demographics benefit from it.

The THP's multimorbidity scores can be distinguished from traditional multimorbidity scores in three ways: First, they are derived from a comprehensive set of diagnostic, prescription, lab, and medical procedure data. This is in contrast to other multimorbidity scores that use only one set of information, such as diagnoses (as is the case with the ECI [10] and the CCI [9]) or prescriptions (such as Rx-Risk [29]). Second, these scores were derived from a large and diverse cohort of 794,294 patients with medical data spanning decades. Third, the THP was calculated from patients of both sexes and from across the age spectrum (3-90 years), rather than focusing on mostly geriatric populations as with traditional multimorbidity. To the best of our knowledge, this is the first time that ML was used to integrate multiple types of physician-curated clinical information from a large, diverse population and produce a multimorbidity score that can help guide patient care irrespective of sex and age.

As part of the overall multimorbidity score in the THP, we calculated robust, organ-system-specific scores that provide a more granular picture of health. We believe that these organ-specific multimorbidity scores can complement existing condition-specific scores in clinical practice by providing additional validation for treatment decisions for cardiovascular, respiratory, neuropsychiatric, gastrointestinal, and renal domains. Along these lines, we note that these disease-specific scores often use patient-reported outcomes as part of their input [14], with some even using them exclusively [3]. Although EHR software systems may have health-based modules to automatically compute such scores at the population level, these self-reported data are frequently unavailable or unreliable [30], making it difficult to scale these scores to the population level with a high degree of efficacy. Although the THP cannot be directly compared against these alternative risk scores, as they typically focus on diagnoses versus emergency events, the fact that the THP consistently achieved relatively high AUROCs is nevertheless promising with regard to its ability to complement these more specific risk scores. Specifically, it says something well established in multimorbidity scores but understudied in more specific risk scores: foregoing patient input (which typically contains useful information) entirely, in exchange for more scalable data, can still lead to strong results. Moreover, these alternate risk scores are also typically hyperspecific, limiting their clinical utility to a subset of patients—likely due to them being built on similarly restricted cohorts (eg the American Heart Association pooled cohort equations for atherosclerotic cardiovascular disease derived from cohorts exclusively in the age range of 40-79 years). As our approach has no constraints upon individual patients' age or sex, and are built using a similarly diverse cohort, risk profiles that are applicable to a far larger population can be easily derived. Of course, we assessed generalizability only among three well-known dimensions (age, sex, and income), and there are far more subtle biases that have been observed even among established risk scores, such as the CHADS2VASC stroke score underestimating risk in patients with chronic renal disease [31]. Further study will be needed to fully examine these sorts of biases in our proposed risk models, but even in this case, the scalability of our approach will only make this research simpler to perform.

Limitations

Data-related limitations of this study include unmeasured variables and incomplete observations. Regarding the former, in this study, we did not include lifestyle behavioral data, such as nutrition, smoking, and physical activity. Although reporting of these factors is known to be inconsistent and unreliable [32], especially in healthy populations (which typically lack recent EHR/claims medical history), they play an important role in clinical outcomes. We believe this would be most addressable through the collection of passive data from wearable sensors, which future work will include. On a similar note, although we were able to use aggregate statistics for race and economic status based on zip-code-derived census data, we were unable to track them at an individual level. Though this form of zip code aggregation has been shown to be useful in clinical risk assessment [33], individual SDoH data could increase the precision and accuracy of THP multimorbidity scores. Future studies of the THP will examine the impact of longer observation and follow-up windows on strategies for clinical intervention. Finally, we note the unreliability of claims data at large, as they are typically produced with financial incentives that are not necessarily aligned with patient care, though they are still often used for risk assessment problems [34,35].

Conclusion

In summary, we combined practical clinical knowledge with modern ML on large-scale data to produce THP multimorbidity scores to aid in decision making across generalizable patient populations. We believe that the THP will allow for more targeted prioritization of care-gap closure, the assessment of comprehensive risk profiles for a greater number of patients, and facilitation of better physician-patient interactions and joint decision making via feature explainability. Although prospective studies will be required to measure the utility of this approach, our intention is that the THS may be used as a preliminary risk stratifier to rapidly prioritize patients for care from a population health management perspective [36]. Once a patient is engaged with a care provider, the organ-specific scores can be used to guide, and explain, individualized clinical interventions based on existing best practices. This would provide the foundation for an integrated continuum between population health and personalized medicine. Finally, we also note the promise that the THP has for clinical research at large, reflecting the rare opportunity to study holistic clinical risk at an extreme scale, potentially unveiling clinically valuable insights.

Acknowledgments

We gratefully acknowledge Chirag J Patel, Ashley Dumm, Will Stedden, and Sirinda Wongpanich for helpful suggestions during the study, comments on the draft, and minor revisions. The data are not publicly released due to Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations and patient privacy. The corresponding author may be contacted to discuss data access following approved agreements. The source code for model creation/validation and figure/table generation will be made publicly available through a GitHub repository upon publication.

Authors' Contributions

BN conceived of the research idea, with critical input from AM, AD, and GL. AM created the data ingestion pipeline/modeling infrastructure source code. AM and AB performed the statistical analysis and created all figures/tables. GL designed the analytical framework and provided oversight for analysis. AD developed the clinical features and labels. GL and AD created the list of medical codes used in the features/labels and provided health care domain expertise. AD collected all citations/references used in this study. AM drafted the initial manuscript, with heavy critical revision by AD and GL. BN provided feedback and editing. BN and AM had access to the underlying data and verified them. All the authors gave final approval for the completed manuscript version.

Conflicts of Interest

All research was funded by Anthem AI, an R&D group of Anthem Inc. AM, AB, and BN are full-time employees at Anthem AI. AD and GL are full-time and part-time employees, respectively, at XY.ai. All investigators received no compensation outside of their regular compensation of their respective employers. AM and BN report a patent pending on the scores included in the total health profile (THP).

References

1. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. Ann Fam Med 2012;10(2):134-141 [FREE Full text] [doi: 10.1370/afm.1363] [Medline: 22412005]
2. Soh CH, Ul Hassan SW, Sacre J, Maier AB. Morbidity measures predicting mortality in inpatients: a systematic review. J Am Med Dir Assoc 2020 Apr;21(4):462-468.e7. [doi: 10.1016/j.jamda.2019.12.001] [Medline: 31948852]
3. Newman AB, Boudreau RM, Naydeck BL, Fried LF, Harris TB. A physiologic index of comorbidity: relationship to mortality and disability. J Gerontol A Biol Sci Med Sci 2008 Jun;63(6):603-609 [FREE Full text] [doi: 10.1093/gerona/63.6.603] [Medline: 18559635]
4. Fortin M, Bravo G, Hudon C, Lapointe L, Almirall J, Dubois M, et al. Relationship between multimorbidity and health-related quality of life of patients in primary care. Qual Life Res 2006 Feb;15(1):83-91. [doi: 10.1007/s11136-005-8661-z] [Medline: 16411033]
5. Librero J, Peiró S, Ordiñana R. Chronic comorbidity and outcomes of hospital care. J Clin Epidemiol 1999 Mar;52(3):171-179. [doi: 10.1016/s0895-4356(98)00160-7] [Medline: 10210233]
6. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med 2002 Nov 11;162(20):2269-2276. [doi: 10.1001/archinte.162.20.2269] [Medline: 12418941]
7. Marengoni A, von Strauss E, Rizzuto D, Winblad B, Fratiglioni L. The impact of chronic multimorbidity and disability on functional decline and survival in elderly persons. A community-based, longitudinal study. J Intern Med 2009 Feb;265(2):288-295 [FREE Full text] [doi: 10.1111/j.1365-2796.2008.02017.x] [Medline: 19192038]
8. Payne RA, Mendonca SC, Elliott MN, Saunders CL, Edwards DA, Marshall M, et al. Development and validation of the Cambridge Multimorbidity Score. CMAJ 2020 Feb 03;192(5):E107-E114 [FREE Full text] [doi: 10.1503/cmaj.190757] [Medline: 32015079]
9. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
10. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009 Jun;47(6):626-633. [doi: 10.1097/MLR.0b013e31819432e5] [Medline: 19433995]
11. Stirland LE, González-Saavedra L, Mullin DS, Ritchie CW, Muniz-Terrera G, Russ TC. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. BMJ 2020 Feb 18;368:m160 [FREE Full text] [doi: 10.1136/bmj.m160] [Medline: 32071114]
12. King DE, Xiang J, Pilkerton CS. Multimorbidity trends in United States Adults, 1988-2014. J Am Board Fam Med 2018;31(4):503-513 [FREE Full text] [doi: 10.3122/jabfm.2018.04.180008] [Medline: 29986975]
13. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract 2018 Apr;68(669):e245-e251 [FREE Full text] [doi: 10.3399/bjgp18X695465] [Medline: 29530918]

14. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998 May 12;97(18):1837-1847. [doi: 10.1161/01.cir.97.18.1837] [Medline: 9603539]
15. Marengoni A, Onder G. Guidelines, polypharmacy, and drug-drug interactions in patients with multimorbidity. BMJ 2015 Mar 11;350:h1059. [doi: 10.1136/bmj.h1059] [Medline: 25761379]
16. Salisbury C. Multimorbidity: time for action rather than words. Br J Gen Pract 2013 Feb;63(607):64-65 [FREE Full text] [doi: 10.3399/bjgp13X661020] [Medline: 23561658]
17. Alzoubi H, Alzubi R, Ramzan N, West D, Al-Hadhrami T, Alazab M. A review of automatic phenotyping approaches using electronic health records. Electronics 2019 Oct 29;8(11):1235 [FREE Full text] [doi: 10.3390/electronics8111235]
18. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc 2013 Dec;20(e2):e206-e211 [FREE Full text] [doi: 10.1136/amiajnl-2013-002428] [Medline: 24302669]
19. Hassaine A, Salimi-Khorshidi G, Canoy D, Rahimi K. Untangling the complexity of multimorbidity with machine learning. Mech Ageing Dev 2020 Sep;190:111325 [FREE Full text] [doi: 10.1016/j.mad.2020.111325] [Medline: 32768443]
20. Xu J, Murphy S, Kochanek K, Arias E. Mortality in the United States, 2018: NCHS data brief No. 355, January 2020. URL: https://www.cdc.gov/nchs/products/databriefs/db355.htm [accessed 2021-07-01]
21. Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. Int J Med Inform 2018 Aug;116:10-17 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.05.006] [Medline: 29887230]
22. U.S. Preventive Services Task Force. Recommendations search results. URL: https://www.uspreventiveservicestaskforce.org/uspstf/topic_search_results?topic_status=All&searchterm= [accessed 2021-07-01]
23. Norgeot B, Quer G, Beaulieu-Jones B, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020 Sep;26(9):1320-1324 [FREE Full text] [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]
24. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [doi: 10.1148/radiology.143.1.7063747] [Medline: 7063747]
25. Jia H, Lubetkin EI. Impact of nine chronic conditions for US adults aged 65 years and older: an application of a hybrid estimator of quality-adjusted life years throughout remainder of lifetime. Qual Life Res 2016 Aug 18;25(8):1921-1929. [doi: 10.1007/s11136-016-1226-5] [Medline: 26781442]
26. Ryan A, Wallace E, O'Hara P, Smith SM. Multimorbidity and functional decline in community-dwelling adults: a systematic review. Health Qual Life Outcomes 2015 Oct 15;13:168 [FREE Full text] [doi: 10.1186/s12955-015-0355-9] [Medline: 26467295]
27. McPhail SM. Multimorbidity in chronic disease: impact on health care resources and costs. Risk Manag Healthc Policy 2016;9:143-156 [FREE Full text] [doi: 10.2147/RMHP.S97248] [Medline: 27462182]
28. Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. Health Qual Life Outcomes 2009 Sep 08;7:82 [FREE Full text] [doi: 10.1186/1477-7525-7-82] [Medline: 19737412]
29. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992 Feb;45(2):197-203. [doi: 10.1016/0895-4356(92)90016-g] [Medline: 1573438]
30. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
31. Foote C, Woodward M, Jardine MJ. Scoring risk scores: considerations before incorporating clinical risk prediction tools into your practice. Am J Kidney Dis 2017 May;69(5):555-557. [doi: 10.1053/j.ajkd.2017.02.005] [Medline: 28434521]
32. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990 Jan;43(1):87-91. [doi: 10.1016/0895-4356(90)90060-3] [Medline: 2319285]
33. Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. J Am Med Inform Assoc 2018 Jan 01;25(1):47-53 [FREE Full text] [doi: 10.1093/jamia/ocx130] [Medline: 29177457]
34. Patterson-Lomba O, Ayyagari R, Carroll B. Risk assessment and prediction of TD incidence in psychiatric patients taking concomitant antipsychotics: a retrospective data analysis. BMC Neurol 2019 Jul 20;19(1):174 [FREE Full text] [doi: 10.1186/s12883-019-1385-4] [Medline: 31325958]
35. Martin J, Mills S, Foley ME. Innovative models of dental care delivery and coverage: patient-centric dental benefits based on digital oral health risk assessment. Dent Clin North Am 2018 Apr;62(2):319-325. [doi: 10.1016/j.cden.2017.11.009] [Medline: 29478460]
36. Haas LR, Takahashi PY, Shah ND, Stroebel RJ, Bernard ME, Finnie DM, et al. Risk-stratification methods for identifying patients for care coordination. Am J Manag Care 2013 Sep;19(9):725-732 [FREE Full text] [Medline: 24304255]

Abbreviations

AHRQ: Agency for Health care Research and Quality

AUC: area under the curve
AUROC: area under the receiver operating curve
CCI: Charlson Comorbidity Index
CPT: Current Procedural Terminology
dx: diagnosis
ECI: Elixhauser Comorbidity Index
eGFR: estimated glomerular filtration rate
EHR: electronic health record
GPI: General Product Identifier
HbA_{1c}: hemoglobin A_{1c}
HCUP: Healthcare Cost and Utilization Project
HIPAA: Health Insurance Portability and Accountability Act of 1996
ICD-10: International Classification of Diseases, Tenth Revision
IP: inpatient
MI-CLAIM: Minimum Information about Clinical Artificial Intelligence Modeling
ML: machine learning
SDoH: social determinants of health
THP: total health profile
THS: total health score

Edited by G Eysenbach; submitted 13.08.21; peer-reviewed by X Cheng, J Walsh; comments to author 03.09.21; revised version received 15.09.21; accepted 18.09.21; published 26.11.21.

Please cite as:
Mahajan A, Deonarine A, Bernal A, Lyons G, Norgeot B
Developing the Total Health Profile, a Generalizable Unified Set of Multimorbidity Risk Scores Derived From Machine Learning for Broad Patient Populations: Retrospective Cohort Study
J Med Internet Res 2021;23(11):e32900
URL: https://www.jmir.org/2021/11/e32900
doi:10.2196/32900
PMID:34842542

Original Paper

Mobile Diary App Versus Paper-Based Diary Cards for Patients With Borderline Personality Disorder: Economic Evaluation

Sidsel Lund Laursen[1], BA, MSc; Stig Helweg-Jørgensen[2], BA, MSc, PhD; Astrid Langergaard[1], BA, MSc; Jesper Søndergaard[1], BA, MSc; Sabrina Storgaard Sørensen[1], BA, MSc, PhD; Kim Mathiasen[2,3], BA, MSc, PhD; Mia Beck Lichtenstein[2,3], BA, MSc, PhD; Lars Holger Ehlers[1], BA, MSc, PhD

[1]Department of Clinical Medicine, Danish Center for Healthcare Improvements, Aalborg University, Aalborg Oest, Denmark

[2]Research Unit for Telepsychiatry and E-mental Health, Mental Health Services in the Region of Southern Denmark, Odense, Denmark

[3]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Lars Holger Ehlers, BA, MSc, PhD
Department of Clinical Medicine
Danish Center for Healthcare Improvements
Aalborg University
Fredrik Bajers Vej 5
Aalborg Oest, 9220
Denmark
Phone: 45 99408257
Email: lehlers@dcm.aau.dk

Abstract

Background: The cost-effectiveness of using a mobile diary app as an adjunct in dialectical behavior therapy (DBT) in patients with borderline personality disorder is unknown.

Objective: This study aims to perform an economic evaluation of a mobile diary app compared with paper-based diary cards in DBT treatment for patients with borderline personality disorder in a psychiatric outpatient facility.

Methods: This study was conducted alongside a pragmatic, multicenter, randomized controlled trial. The participants were recruited at 5 Danish psychiatric outpatient facilities and were randomized to register the emotions, urges, and skills used in a mobile diary app or on paper-based diary cards. The participants in both groups received DBT delivered by the therapists. A cost-consequence analysis with a time horizon of 12 months was performed. Consequences included quality-adjusted life years (QALYs), depression severity, borderline severity, suicidal behavior, health care use, treatment compliance, and system usability. All relevant costs were included. Focus group interviews were conducted with patients, therapists, researchers, and industry representatives to discuss the potential advantages and disadvantages of using a mobile diary app.

Results: A total of 78 participants were included in the analysis. An insignificantly higher number of participants in the paper group dropped out before the start of treatment (P=.07). Of those starting treatment, participants in the app group had an average of 37.1 (SE 27.55) more days of treatment and recorded an average of 3.16 (SE 5.10) more skills per week than participants in the paper group. Participants in both groups had a QALY gain and a decrease in depression severity, borderline severity, and suicidal behavior. Significant differences were found in favor of the paper group for both QALY gain (adjusted difference −0.054; SE 0.03) and reduction in depression severity (adjusted difference −1.11; SE 1.57). The between-group difference in total costs ranged from US $107.37 to US $322.10 per participant during the 12 months. The use of services in the health care sector was similar across both time points and groups (difference: psychiatric hospitalization <5 and <5; general practice −1.32; SE 3.68 and 2.02; SE 3.19). Overall, the patients showed high acceptability and considered the app as being easy to use. Therapists worried about potential negative influences on the therapist-patient interaction from new work tasks accompanying the introduction of the new technology but pointed at innovation potential from digital database registrations.

Conclusions: This study suggests both positive and negative consequences of mobile diary apps as adjuncts to DBT compared with paper diary cards. More research is needed to draw conclusions regarding its cost-effectiveness.

Trial Registration: ClinicalTrials.gov NCT03191565; http://clinicaltrials.gov/ct2/show/NCT03191565

International Registered Report Identifier (IRRID): RR2-10.2196/17737

(J Med Internet Res 2021;23(11):e28874) doi:10.2196/28874

KEYWORDS

borderline personality disorder; dialectical behavior therapy; mobile app; psychotherapy; cost-consequence; mHealth; mobile phone

Introduction

Background

Borderline personality disorder (BPD) is estimated to affect between 0.7% and 4.8% of the general population in Scandinavia and accounts for approximately 15% of adult admissions in Danish psychiatric hospitals [1-3]. Patients diagnosed with BPD are often characterized as emotionally unstable, impulsive, and self-harming and have unstable relationships [4]. It has been shown that the mortality rate is 8 times higher in patients with BPD than in the general population [5]. Furthermore, BPD has large societal consequences, as one study has found that the total mean societal cost related to BPD is US $32,863.01 per patient per year [6].

One of the most well-researched and clinically effective psychosocial treatments for patients with BPD is dialectical behavior therapy (DBT) [7-12]. Studies have shown DBT to be effective in reducing nonsuicidal self-injury and increasing quality-adjusted life years (QALYs), when compared with treatment as usual or client-centered therapy [13-16]. The DBT includes elements from cognitive-behavioral therapy and is centered on learning a set of predefined skills, such as mindfulness, tolerance of distress, regulation of emotions, and navigation of interpersonal situations [17,18]. Learning the DBT skills is done through self-monitoring using weekly paper-based diary cards filled out by the patient [19,20]. The use of paper-based diary cards, however, has limitations for both patients and therapists. Working with paper diaries is burdensome and may reduce patient engagement in treatment [21,22]. Retrieving and reviewing paper diaries to evaluate the patient's progress over time is a time-consuming process for the therapist, and it is not possible for the therapist to follow the patient alongside self-treatment at home [20-22].

New studies suggest that mobile diary apps can overcome some of these limitations and be more effective in reducing distress symptoms and suicidal and nonsuicidal self-harming behaviors compared with treatment with DBT alone [23-25]. Mobile apps have the potential to reduce barriers and increase engagement in treatment [21,22]. Mobile apps can automatically provide an overview of the patients' progress and enable therapists to monitor the present emotional stage of patients during treatment [21,24,26].

Objective

Economic evaluations of new health technology constitute important knowledge for decision-makers in health care as they estimate the value for money of the technology compared with alternative use of the scarce financial resources. To our knowledge, no economic evaluation has yet been conducted on the use of mobile apps as adjuncts to DBT in patients with BPD. Therefore, the aim of this economic evaluation is to investigate the costs and consequences of a mobile diary app compared with paper-based diary cards for patients with BPD, within the context of a psychiatric outpatient facility.

Methods

Overview

This economic evaluation was reported in line with the Consolidated Health Economic Evaluation Reporting Standards [27]. It was conducted alongside a pragmatic, multicenter, randomized controlled trial as part of the ENTER program [28]. The clinical study is described in detail in the study protocol [29] but will be summarized in the Methods section of this evaluation. Economic evaluation was carried out as a cost-consequence analysis, which is why costs and consequences are shown separately [30]. The cost-consequence analysis approach was chosen because of the possibility of investigating a broad range of consequences [30], such as intermediate outcomes and qualitative aspects. The time horizon was restricted to 12 months, and all relevant costs were included in the analysis. All costs were measured in Danish Krone (DKK) and subsequently converted to the US dollars (US $).

Participants

The participants were recruited from 5 Danish psychiatric outpatient facilities between June 2017 and December 2018. The participants were eligible for inclusion if they were ≥18 years of age, formally diagnosed by a psychiatric specialist with Emotionally Unstable Personality Disorder (F60.3 according to ICD-10 criteria) [31], and admitted for psychiatric outpatient treatment. To be included, participants also had to be either self-harming or to have had suicidal behavior within the last 3 years. Participants with comorbid disorders, such as depression, anxiety, and posttraumatic stress disorder, were allowed to participate. On the contrary, participants were excluded if they had no access to a working smartphone; an IQ <70; or a comorbid disorder such as substance abuse, bipolar disorder, or a disorder within the schizophrenic spectrum. Before entering the study, all participants signed an informed consent form.

Interventions

Participants were randomly assigned to either the mobile diary app group or to the paper-based diary card group using REDCap (Research Electronic Data Capture) [29]. In both groups, participants had to register entries about aspects such as emotional dysregulation, suicidal and self-harm thoughts, and skill use. The participants received the paper-based treatment registered daily on paper diary cards, whereas the participants using the mobile diary app were registered daily via the mDiary app (The Monsenso system). The mDiary app is a market solution containing psychoeducational material and visualizations of the participant's data, which can be used in therapy sessions and for real time monitoring alongside treatment [32]. Participants in both groups received standard

DBT (delivered as individual sessions, group sessions, and telephone coaching). The therapists instructed the participants on how to use the app or paper diary cards. Other treatments, such as medications, were allowed. Passive sensor data (eg, activity level and phone use) were collected from all participants through their smartphones. The treatment period was 40 weeks at 2 sites and 12 months at the other sites, and neither the participants nor the therapists were blinded to the group allocation.

Consequences

Health-Related Outcomes

Health-related quality of life was based on the EuroQol 5-Dimensions 5-Levels (EQ-5D-5L) questionnaire [33]; depression severity from the Patient Health Questionnaire-9 [34]; borderline severity was assessed using the Zanarini Rating Scale for BPD [35]; and suicide behavior from the Suicide Behaviors Questionnaire [36]. Data were collected at baseline and at the 12-month follow-up. The participants' utility scores were calculated using the Danish weights for the EQ-5D-5L questionnaire [37]. QALY gain was estimated by assuming linear interpolation between utility scores at baseline and a 12-month follow-up, after which the area under the curve was calculated [38]. For depression severity, borderline severity, and suicide behavior, the difference from baseline to follow-up was calculated and compared between the groups.

Health Care Sector

In Denmark, the civil registration number makes it possible to link information from several registries to an individual [39]. Information on psychiatric hospitalizations and consultations with general practice was obtained at an individual level from the Primary National Health Insurance Service Register and the Danish National Patient Registry [40,41]. The specific purpose of the consultations in general practice was not available, which is why all consultations were included. The number of psychiatric hospitalizations and consultations with general practice was estimated 30 days before and after baseline and 12 months before and after baseline, respectively.

Treatment Compliance and Skill Recordings

It was continuously assessed whether the participants were still receiving treatment and whether they were recording their skills use. The recording of the skills used per week was estimated at an individual level as the total number of skills for each participant divided by the number of days in treatment, which was subsequently calculated for a week. A participating therapist estimated the time spent on recording in the 2 groups based on experience from the trial.

The usability of the mobile diary app was assessed by patients and therapists using a system usability scale [42]. This assessment was conducted in focus group interviews alongside a feasibility study of the app.

Future innovation potential was assessed by app developers, researchers, and other personnel related to the study. This assessment was conducted during a workshop to discuss the advantages, disadvantages, and development potential.

Costs

All relevant costs were included in the analysis and adjusted to the price level in 2019 using the Danish net price index [43]. The conversion rate from kr to US $ was kr 634.0 per US $100 from September 23, 2021 [44]. The costs were not discounted because of a time horizon of 12 months [45].

The costs of health care services, patient costs, and municipality costs were used to adjust for baseline differences in the analysis. Health care services included primary care services, prescription medicine, inpatient and outpatient services, and emergency contacts to psychiatric or somatic hospitals. Patient costs included out-of-pocket medical expenses, including expenses for both medicine and primary care services. Municipality costs included nurse care, daily care, and domestic help at home. Information on each participant's use of health care services and prices for public services was retrieved from the Danish Primary National Health Insurance Service Register, the Danish National Patient Registry, and the Danish National Prescription Registry [40,41,46]. Information regarding municipality costs was collected based on the Treatment Inventory of Costs in Psychiatric Patients questionnaire [47], which was administered to all participants at baseline. Baseline costs were calculated for resource use 12 months before the start of the study until the baseline date.

Information on intervention costs was provided by the key personnel at a psychiatric outpatient facility. The intervention costs of the app were estimated as the capital and operating costs. Capital costs included the one-time cost of starting and establishing the mobile diary app program and the education of therapists on how to use the program's software. Operating costs included costs, such as program licenses. Startup establishment and program license costs were estimated as a range due to price instability of new devices, especially if the device is new in the market [48,49]. The pricing was estimated in collaboration with the software company and public purchasers, with an expected discount if the software were to be implemented in Denmark. The license fee was allocated to an estimated number of patients per therapist. The costs of educating the therapists were estimated using the time allocated for training and the average effective hourly wages for the therapists. The wages were based on information from Statistics Denmark on the national average gross wages corresponding to the therapists' occupation [50]. Capital costs were annuitized over 3 years, with a discount rate of 4% per year [51,52]. Subsequently, the costs were allocated to an estimated number of patients per therapist per year and were included in the analysis as annual costs. The relevant intervention costs in the paper group included the price of paper and the printing of diaries for 12 months. The choice between the app and paper was assumed to not influence the overhead costs. See Multimedia Appendix 1 [43,44,48-53] for further information on cost estimation.

Statistical Analysis

To account for missing data, multiple imputation was performed, which was applicable because the missing data were assumed to be missing at random [54,55]. The consequences of borderline severity and suicide behavior are shown as nonimputed estimates

due to a high number of missing values at both baseline (17 and 20 missing values, respectively) and follow-up (20 and 17 missing values, respectively). The analyses were conducted using an intention-to-treat analysis. See Multimedia Appendix 2 [54] for a summary of the missing data and a full description of the imputation model.

Data are reported either as the number of participants in each group and percentages, median, and IQR, or as means and SE. Incremental health-related outcomes were estimated as both unadjusted and adjusted outcomes based on the regression analysis. The health-related outcomes were adjusted for age, sex, relationship status, education level, baseline scores (utility score, depression score, suicide score, and borderline score), and baseline costs to control for baseline differences [38]. For consequences related to the health care sector and treatment compliance, *P* values were calculated using the Fisher exact test for categorical variables and Student *t* test (2-tailed) for continuous variables [56]. *Statistical* significance was set at a *P* value <.05. Statistical analyses were performed using the STATA (version 16.1, StataCorp).

Ethics and Consent to Participate

The trial was registered at ClinicalTrials.gov (identifier NCT03191565) and was performed in accordance with the Declaration of Helsinki and approved by the Danish Ethics Committee in the Region of Southern Denmark (Registration number S-20160085). Participants were recruited between June 2017 and December 2018. All participants signed an informed consent form before participating in the study.

Data Sharing Statements

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Results

Overview

In total, 79 participants were enrolled in the trial, of whom 78 were eligible for inclusion in the analysis. Participants in both groups were predominantly women, with a mean age of 29 years (Table 1).

Table 1. Characteristics of the participants at baseline.

Study population	Mobile diary app (N=42)	Paper diary cards (N=36)
Age (years), median (IQR)	26.4 (22.2-32.8)	27.2 (22.3-34.6)
Sex (female), n (%)	>37 (≥88)	>31 (≥86)
Cohabiting status, n (%)		
Living with somebody	14 (33)	18 (50)
Living alone	28 (67)	18 (50)
Education, n (%)		
Education <3 years[a]	>37 (≥88)	27 (75)
Education ≥3 years[b]	<5 (≤12)	9 (25)
Employment, n (%)		
In employment	9 (21)	9 (25)
Not in employment	33 (79)	27 (75)
Costs[c] (US $), mean (SE)		
Health care costs	17,200.35 (3351.31)	15,684.14 (2800.31)
Municipality costs	7785.36 (5680.32)	250.08 (140.15)
Patient costs	303.66 (49.47)	327.02 (48.09)
Total costs	25,289.37 (6918.34)	16,378.66 (2805.81)

[a]Participants with highest education of public school, high school, further education <3 years and apprenticeship.

[b]Participants with further education ≥3 years.

[c]Annual cost the year before the study start date.

Consequences

Participants in both groups experienced an increase in utility scores from baseline to the 12-month follow-up (Table 2). The adjusted QALY gain difference between the app group and the paper group was −0.054 (SE 0.03; *P*<.001), indicating a smaller QALY gain in the app group. The unadjusted QALY gain difference was close to the adjusted estimate. Both groups showed improvements in the EQ-5D domains, except for domain 1 (mobility) in the app group and domain 2 (self-care) in the paper group. The improvement in domain 5 (anxiety and depression) was comparable between the 2 groups.

Table 2. Quantitative consequences of the mobile diary app and paper-based diary cards.

Health-related outcomes	Mobile diary app (N=42)	Paper diary cards (N=36)	Unadjusted between-group difference	*P* value	Adjusted between-group difference	*P* value
EQ-5D[a] levels, utility and QALY[b], mean (SE)[c]						
Baseline			—[d]	—	—	—
Domain 1- Mobility	1.86 (0.15)	1.89 (0.17)				
Domain 2—Self-care	1.52 (0.14)	1.31 (0.10)				
Domain 3—Usual activities	2.86 (0.14)	2.67 (0.15)				
Domain 4—Pain or discomfort	2.93 (0.16)	2.58 (0.21)				
Domain 5—Anxiety or depression	3.17 (0.17)	3.11 (0.15)				
Utility score	0.45 (0.05)	0.51 (0.05)				
Follow-up			—	—	—	—
Domain 1-Mobility	2.2 (0.22)	1.57 (0.13)				
Domain 2-Self-care	1.37 (0.13)	1.41 (0.14)				
Domain 3-Usual activities	2.56 (0.19)	1.95 (0.17)				
Domain 4-Pain or discomfort	2.66 (0.17)	2.30 (0.15)				
Domain 5-Anxiety or depression	2.41 (0.17)	2.33 (0.18)				
Utility score	0.6 (0.06)	0.74 (0.03)				
QALY gain	0.078 (0.03)	0.115 (0.03)	-0.037 (0.04)	.38	-0.054 (0.03)	<.001
Borderline severity, mean (SE)[e]						
Baseline	18.6 (1.22)	19.5 (1.08)	—		—	—
Follow-up	8.82 (1.16)	8.83 (1.45)	—		—	—
Change within the group	9.85 (1.28)	10.67 (1.39)	−0.81 (1.93)	.67	−0.67 (1.73)	.004
Depression severity, mean (SE)[c]						
Baseline	17.52 (0.80)	16.11 (0.88)	—	—	—	—
Follow-up	12.34 (1.03)	10.47 (1.3)	—	—	—	—
Change within the group	5.18 (1.01)	5.64 (1.38)	−0.46 (1.67)	.78	−1.11 (1.57)	.04
Suicidal behavior, mean (SE)[e]						
Baseline	11.60 (0.75)	11.96 (0.88)	—	—	—	—
Follow-up	8.86 (0.58)	9.38 (0.62)	—	—	—	—
Change within the group	2.79 (0.59)	2.76 (0.72)	0.03 (0.93)	.98	0.47 (0.73)	<.001

[a]EQ-5D: EuroQol 5-Dimensions.

[b]QALY: quality-adjusted life year.

[c]Imputed data set.

[d]Empty cells concern descriptive measures not relevant for statistical testing.

[e]Variables with missing values.

Participants in both groups had decreased borderline severity, depressive severity, and suicidal behavior during the 12-month follow-up. The adjusted difference in the change within the group in depression severity between the app group and the paper group was −1.11 (SE 1.57; *P*=.04), indicating that participants in the app group had a smaller decrease in depression severity. The unadjusted differences showed a similar result. In addition, participants in the app group had a smaller decrease in borderline severity but a greater decrease in suicidal behavior. These results should be interpreted with reservations due to a high number of missing values at both baseline and at follow-up.

The use of health care sector services was similar across both the periods and the groups (difference: psychiatric hospitalization <5 and <5; general practice −1.32, SE 3.68 and 2.02, SE 3.19), indicating that registering in the app compared with on paper does not affect the resource use in the health care sector (Table 3).

The paper group had an insignificantly higher number of participants dropping out before the start of treatment compared with the app group (*P*=.07). Of those starting treatment, participants in the app group were more persistent in adhering to treatment and had an average of 37.1 (SE 27.55) more days of treatment than the paper group. Participants in the app group registered 3.16 (SE 5.10) more skills per week compared with participants in the paper group. Therapists and researchers estimated that the app was time saving for the patient, approximately 1 minute on each diary entry. They further pointed out a potential utility gain for patients from the possibility of choosing according to preferences between registering in an app or on paper diary cards.

Table 3. Quantitative consequences of the mobile diary app and paper-based diary cards.

Health-related outcomes	Mobile diary app (n=42)	Paper diary cards (n=36)	Between-group difference	*P* value
Health care sector contacts				
Participants hospitalized in a psychiatric hospital, n (%)				
30 days before baseline	<5 (≤12)	<5 (≤14)	≤5	>.99
30 days after baseline	<5 (≤12)	<5 (≤14)	≤5	>.99
General practice contacts, mean (SE)				
12 months before baseline	19.42 (2.4)	20.75 (2.83)	−1.32 (3.68)	.72
12 months after baseline	19.02 (2.18)	17 (2.32)	2.02 (3.19)	.53
Treatment compliance				
Participants, who never start treatment, n (%)	<5 (≤12)	7 (19)	≤−2	.07
Number of treatment days for participants starting treatment, mean (SE)	265.93 (18.64)	228.83 (19.64)	37.1 (27.55)	.18
Skill recordings, mean (SE)				
All participants				
Skills recorded per patient	647.93 (150.72)	409.89 (173.99)	238.04 (229)	.30
Participants starting treatment				
Skills recorded per patient	680.33 (156.56)	508.83 (212.54)	171.5 (257.9)	.51
Per week	15.04 (3.04)	11.88 (4.29)	3.16 (5.10)	.54
Time spent on recording				
Time spent on completing the diary (minutes)	4	5	−1	—[a]
System Usability Scale, mean (SD)				
Patients (n=16)	81.2 (9.9)	—	—	—
Therapists (n=23)	61.6 (18.6)	—	—	—

[a]System usability scale is only reported for the mobile diary group.

The System Usability Score was 81.2 (SD 9.9) in patients, but significantly lower at 61.6 (SD 18.6) in therapists (for further information see Marceau et al [22]). In general, patients were satisfied with the mobile solution, whereas the therapists expressed concern regarding the potential for a negative influence on the patient-therapist interaction due to new tasks related to the implementation of the new technology [22]. In addition, reduced flexibility in data collection due to the standard structure of the app was described as a potential disadvantage. Finally, numerous implementation issues were highlighted in the focus groups, including concerns relating to a potential lack of technical skills of both the patient and therapist, and security and storage of the collected data.

Future potential for innovation was considered relevant as electronic data recording can be used to predict outcomes and benefits from changes in treatment strategy (Textbox 1). The mobile diary app could provide therapists with information on the patients alongside treatment, thereby making it possible to define cut-off values for intervening outside sessions. Program optimization and visualization of the entered data could further inform the patients about their progress, potentially increasing the patient's self-insight and empowerment.

Textbox 1. Innovation potential of the mobile diary app and paper-based diary cards.

Innovation potential

- Real time recording makes it possible to track patients' symptoms and possibly intervene outside sessions.
- Electronic data from the app can potentially be used for predicting outcomes from treatment.

Costs

The between-group difference in total costs during the 12-months ranged from US $107.38 to US $322.13 per participant (Table 4).

Table 4. Intervention costs (US $) per participant at 12-month follow-up.

Intervention cost	Mobile diary app (US $)	Paper diary cards (US $)	Difference (US $)
Startup cost[a]	9.41-28.44	0	9.41-28.24
License	97.96-293.88	0	97.96-293.88
Education	3.27	0	3.27
Paper and print	0	3.26	3.26
Total	110.64-325.39	3.26	107.38-322.13

[a]Annuitized over 3 years with a 4% discount rate.

Discussion

Principal Findings

To our knowledge, this study is the first to evaluate the costs and consequences of a mobile diary app for patients with BPD. Our findings suggest that using a mobile app as an adjunct to DBT comes with a high acceptance of patients, and the potential for increased treatment adherence and future innovation from electronic database registrations. We also found, however, that the app was associated with a reduced improvement in patients' depression and health-related quality of life scores within the first year of treatment. The app led to higher costs but was still a relatively inexpensive intervention.

The economic evaluation was conducted using a cost-consequence approach rather than the traditional cost-utility (ie, cost-per-QALY gained) framework. The cost-consequence approach has been criticized by health economists for not being a normative analysis, as the decision-maker has to choose from her or his own opinion concerning the relative importance of the list of consequences [30]. As the mobile diary app was still in early development, we believe there is a need to investigate a broad range of potential consequences, including intermediate consequences. Indeed, health professionals should be encouraged to engage in the debate about how to evaluate internet-based interventions in mental health. Health economists have concluded that the economic evaluation of devices in health care, in general, represents a number of methodological challenges; in particular, the use of QALY as an outcome may be too narrow to capture all relevant benefits to patients and health care organizations [47]. Specifically, within mental health, the use of QALY and EQ-5D has also been found to be challenging with regard to instrument sensitivity [57].

Participants in both the app and paper groups had a QALY gain, but the gain was significantly larger in the paper group. This result is difficult to explain as the difference in QALY gain between the groups was caused by a worsening in the mobility domain of the EQ-5D-5L among app users (Table 2). There is a risk that patients have interpreted the EQ-5D mobility domain differently, either as a question about physical restrictions or as limitations on mobility due to anxiety or depression. Nevertheless, the results suggest that using mobile apps as adjuncts to DBT might reduce the improvement in quality of life and depression scores within the first year.

There are several limitations to this study, including the small sample size, short follow-up period, and missing data. The missing data on depression severity and utility domains were handled by performing multiple imputations and were assessed as not having an impact on the results [54]. There was a considerable amount of missing data for both borderline severity and suicide behavior, making it infeasible to run the imputation model for these outcomes. Therefore, borderline severity and suicidal behavior were presented as nonimputed data, potentially indicating that these results could be biased [55]. Previously, a moderate correlation has been shown between the EQ-5D utility scores and measures for borderline severity, supporting our findings of a similar improvement in QALY and borderline severity [58].

Even though participants in the app group recorded more skills, this did not contribute to a larger QALY gain in the app group than in the paper group. One possible explanation is that a measurement of an effect like health-related quality of life might not capture the potential effects of the app [59], which is why including other effects is relevant. In this study, potential effects such as new and better information from the patients are highlighted, which is supported by the findings of Lauritsen [60], who concluded that the use of a mobile electronic diary can improve the quality and accuracy of data entered into the diary. A study by Derks et al [61] found that a mobile diary app

together with a biosensor was effective in increasing self-awareness in patients with BPD. Similar results were found in this study, as key personnel stated that increased self-insight and empowerment might be a potential advantage linked to self-monitoring. Furthermore, both in this study and in prior studies [62,63], a mobile app has been proven to enhance treatment compliance. Thus, if this study had been conducted as a cost-utility, the aforementioned consequences would not have been included, meaning the conclusion might have been quite different.

In this study, concerns regarding the app were primarily from therapists who worried that the app would negatively affect their interaction with the patient. In contrast, findings by Austin et al [64] suggests that a mobile app integrated into DBT was useful for patients with BPD in building an alliance with their therapist and that it promoted collaboration between therapist and patient. However, it is important to state that participants who enrolled in this study and the study by Austin et al [64] might have been advocates for technology since they knew the new and alternative treatment would be a mobile app, which could explain why patients were pleased with the app and why it was primarily therapists who had concerns. Thus, the impact of a mobile diary app on the therapist-patient relationship should be explored in future research.

Conclusions

In conclusion, this study suggests that using a mobile diary app as an adjunct to DBT has positive as well as negative consequences for patients and, on average, will lead to higher costs than paper-based diary cards. The use of a mobile diary app in the treatment of patients with BPD is still a relatively new field of research, and further investigation in this area is essential.

Acknowledgments

The authors would like to thank all the participants, both patients and therapists, at 5 different locations. Furthermore, the authors would like to acknowledge the staff at the Center for Telepsychiatry in the Region of Southern Denmark, University of Southern Denmark, and Monsenso for contributing to the mDiary program.

SLL played a primary role in the economic evaluation. Contributions to the design of the trial and data collection were described in the protocol [29]. SLL, AL, JS, SSS, and LHE developed the analytic plan. SLL and AL analyzed the data. SLL, AL, JS, SSS, and LHE participated in data interpretation. SLL drafted the first version of the manuscript. All authors provided critical feedback, helped shape the analysis and writing of the manuscript, and approved the submission.

This study is part of the ENTER project funded by Innovation Fund Denmark (grant number 5159-00002B). The Innovation Fund Denmark has not been involved in the design of the study or the subsequent handling of data. More information regarding the Innovation Fund Denmark can be found at the website [65].

Conflicts of Interest

The mDiary program was developed and owned by the private sector company Monsenso. SHJ has been employed at a public sector psychiatric ward, and he has no affiliation with Monsenso, nor does any of the other researchers.

Multimedia Appendix 1
Detailed information on cost estimates in the economic analysis.
[DOC File , 86 KB - jmir_v23i11e28874_app1.doc]

Multimedia Appendix 2
A detailed description of the used approach for multiple imputation.
[DOC File , 43 KB - jmir_v23i11e28874_app2.doc]

References

1. Ekselius L, Tillfors M, Furmark T, Fredrikson M. Personality disorders in the general population: DSM-IV and ICD-10 defined prevalence as related to sociodemographic profile. Pers Individ Dif 2001 Jan 19;30(2):311-320. [doi: 10.1016/s0191-8869(00)00048-9]
2. Torgersen S, Kringlen E, Cramer V. The prevalence of personality disorders in a community sample. Arch Gen Psychiatry 2001 Jun;58(6):590-596. [doi: 10.1001/archpsyc.58.6.590] [Medline: 11386989]
3. Pedersen L, Simonsen E. Incidence and prevalence rates of personality disorders in Denmark-a register study. Nord J Psychiatry 2014 Nov;68(8):543-548. [doi: 10.3109/08039488.2014.884630] [Medline: 24520919]
4. National Collaborating Centre for Mental Health (UK). Borderline Personality Disorder: Treatment and Management. Leicester (UK): The British Psychological Society & The Royal College of Psychiatrists; 2009.
5. Kjær JN, Biskin R, Vestergaard C, Munk-J Rgensen P. All-cause mortality of hospital-treated borderline personality disorder: a nationwide cohort study. J Pers Disord 2020 Dec;34(6):723-735. [doi: 10.1521/pedi_2018_32_403] [Medline: 30307824]

6. Wagner T, Fydrich T, Stiglmayr C, Marschall P, Salize H, Renneberg B, et al. Societal cost-of-illness in patients with borderline personality disorder one year before, during and after dialectical behavior therapy in routine outpatient care. Behav Res Ther 2014 Oct;61:12-22. [doi: 10.1016/j.brat.2014.07.004] [Medline: 25113523]
7. Stoffers JM, Völlm BA, Rücker G, Timmer A, Huband N, Lieb K. Psychological therapies for people with borderline personality disorder. Cochrane Database Syst Rev 2012 Aug 15(8):CD005652. [doi: 10.1002/14651858.CD005652.pub2] [Medline: 22895952]
8. Rizvi SL, Steffel LM, Carson-Wong A. An overview of dialectical behavior therapy for professional psychologists. Prof Psychol Res Pr 2013;44(2):73-80. [doi: 10.1037/a0029808]
9. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006 Jul;63(7):757-766. [doi: 10.1001/archpsyc.63.7.757] [Medline: 16818865]
10. Linehan MM, Dimeff LA, Reynolds SK, Comtois KA, Welch SS, Heagerty P, et al. Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. Drug Alcohol Depend 2002 Jun 01;67(1):13-26. [doi: 10.1016/s0376-8716(02)00011-x] [Medline: 12062776]
11. McMain SF, Links PS, Gnam WH, Guimond T, Cardish RJ, Korman L, et al. A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. Am J Psychiatry 2009 Dec;166(12):1365-1374. [doi: 10.1176/appi.ajp.2009.09010039] [Medline: 19755574]
12. Verheul R, Van Den Bosch LM, Koeter MW, De Ridder MA, Stijnen T, Van Den Brink W. Dialectical behaviour therapy for women with borderline personality disorder: 12-month, randomised clinical trial in the Netherlands. Br J Psychiatry 2003 Feb;182:135-140. [doi: 10.1192/bjp.182.2.135] [Medline: 12562741]
13. Brettschneider C, Riedel-Heller S, König HH. A systematic review of economic evaluations of treatments for borderline personality disorder. PLoS One 2014 Sep 29;9(9):e107748 [FREE Full text] [doi: 10.1371/journal.pone.0107748] [Medline: 25265185]
14. Murphy A, Bourke J, Flynn D, Kells M, Joyce M. A cost-effectiveness analysis of dialectical behaviour therapy for treating individuals with borderline personality disorder in the community. Ir J Med Sci 2020 May;189(2):415-423. [doi: 10.1007/s11845-019-02091-8] [Medline: 31482522]
15. Priebe S, Bhatti N, Barnicot K, Bremner S, Gaglia A, Katsakou C, et al. Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: a pragmatic randomised controlled trial. Psychother Psychosom 2012;81(6):356-365 [FREE Full text] [doi: 10.1159/000338897] [Medline: 22964561]
16. Swales M, Hibbs RA, Bryning L, Hastings RP. Health related quality of life for young people receiving dialectical behaviour therapy (DBT): a routine outcome-monitoring pilot. Springerplus 2016 Jul 20;5(1):1137 [FREE Full text] [doi: 10.1186/s40064-016-2826-9] [Medline: 27504235]
17. Chapman AL. Dialectical behavior therapy: current indications and unique elements. Psychiatry (Edgmont) 2006 Sep;3(9):62-68 [FREE Full text] [Medline: 20975829]
18. O'Connell B, Dowling M. Dialectical behaviour therapy (DBT) in the treatment of borderline personality disorder. J Psychiatr Ment Health Nurs 2014 Aug;21(6):518-525. [doi: 10.1111/jpm.12116] [Medline: 24191948]
19. Lindenboim N, Comtois KA, Linehan MM. Skills practice in dialectical behavior therapy for suicidal women meeting criteria for borderline personality disorder. Cogn Behav Pract 2007 May;14(2):147-156. [doi: 10.1016/j.cbpra.2006.10.004]
20. Linehan MM. DBT Skills Training Manual, Second Edition. New York: Guilford Publications; 2015.
21. Helweg-Joergensen S, Schmidt T, Lichtenstein MB, Pedersen SS. Using a mobile diary app in the treatment of borderline personality disorder: mixed methods feasibility study. JMIR Form Res 2019 Sep 30;3(3):e12852 [FREE Full text] [doi: 10.2196/12852] [Medline: 31573910]
22. Marceau L, Link C, Smith L, Carolan S, Jamison R. In-clinic use of electronic pain diaries: barriers of implementation among pain physicians. J Pain Symptom Manage 2010 Sep;40(3):391-404 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.12.021] [Medline: 20580526]
23. Rizvi SL, Hughes CD, Thomas MC. The DBT coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
24. Rodante DE, Kaplan MI, Olivera Fedi R, Gagliesi P, Pascali A, Quintero PS, et al. CALMA, a mobile health application, as an accessory to therapy for reduction of suicidal and non-suicidal self-injured behaviors: a pilot cluster randomized controlled trial. Arch Suicide Res 2020 Oct 18:1-18. [doi: 10.1080/13811118.2020.1834476] [Medline: 33073745]
25. Schroeder J, Suh J, Wilks C, Czerwinski M, Munson S, Fogarty J, et al. Data-driven implications for translating evidence-based psychotherapies into technology-delivered interventions. Int Conf Pervasive Comput Technol Healthc 2020 May;2020:274-287 [FREE Full text] [doi: 10.1145/3421937.3421975] [Medline: 33912357]
26. Lederer N, Grechenig T, Baranyi R. unCUT: bridging the gap from paper diary cards towards mobile electronic monitoring solutions in borderline and self-injury. In: Proceedings of the IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH).: IEEE; 2014 Presented at: IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH); May 14-16, 2014; Rio de Janeiro, Brazil. [doi: 10.1109/segah.2014.7067092]

27. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated health economic evaluation reporting standards (CHEERS) --explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value Health 2013;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
28. The mDIARY Study. ENTER. URL: http://www.ementalhealthresearch.dk/27-mdiary.html [accessed 2020-06-25]
29. Helweg-Jørgensen S, Lichtenstein MB, Fruzzetti AE, Møller Dahl C, Pedersen SS. Daily self-monitoring of symptoms and skills learning in patients with borderline personality disorder through a mobile phone app: protocol for a pragmatic randomized controlled trial. JMIR Res Protoc 2020 May 25;9(5):e17737 [FREE Full text] [doi: 10.2196/17737] [Medline: 32449690]
30. Hartfiel N, Edwards R. Cost–consequence analysis of public health interventions. In: Applied Health Economics for Public Health Practice and Research. Oxford, United Kingdom: Oxford University Press; 2019.
31. International statistical classification of diseases and related health problems 10th revision (ICD-10). World Health Organization. 2019. URL: https://icd.who.int/browse10/2019/en#/F60.3 [accessed 2021-01-21]
32. Monsenso. mDiary App. 2021. URL: https://www.monsenso.com/mhealth-mental-health/individuals/ [accessed 2021-07-16]
33. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013 Sep;22(7):1717-1727 [FREE Full text] [doi: 10.1007/s11136-012-0322-4] [Medline: 23184421]
34. 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]
35. Zanarini MC, Vujanovic AA, Parachini EA, Boulanger JL, Frankenburg FR, Hennen J. Zanarini rating scale for borderline personality disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. J Pers Disord 2003 Jun;17(3):233-242. [doi: 10.1521/pedi.17.3.233.22147] [Medline: 12839102]
36. Osman A, Bagge CL, Gutierrez PM, Konick LC, Kopper BA, Barrios FX. The suicidal behaviors questionnaire-revised (SBQ-R): validation with clinical and nonclinical samples. Assessment 2001 Dec;8(4):443-454. [doi: 10.1177/107319110100800409] [Medline: 11785588]
37. Jensen C, Sørensen SS, Gudex C, Jensen M, Pedersen K, Ehlers L. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. Appl Health Econ Health Policy 2021 Jul;19(4):579-591 [FREE Full text] [doi: 10.1007/s40258-021-00639-3] [Medline: 33527304]
38. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005 May;14(5):487-496. [doi: 10.1002/hec.944] [Medline: 15497198]
39. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 2014 Aug;29(8):541-549. [doi: 10.1007/s10654-014-9930-3] [Medline: 24965263]
40. Andersen JS, Olivarius ND, Krasnik A. The Danish National Health Service Register. Scand J Public Health 2011 Jul;39(7 Suppl):34-37. [doi: 10.1177/1403494810394718] [Medline: 21775348]
41. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015 Nov 17;7:449-490 [FREE Full text] [doi: 10.2147/CLEP.S91125] [Medline: 26604824]
42. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
43. Net price index. Statistics Denmark. 2020. URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/net-price-index [accessed 2020-11-04]
44. Valutakurser [Exchange rates]. EUROinvestor. 2019. URL: https://www.valutakurser.dk/ [accessed 2020-06-25]
45. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes. 4th Edition. Oxford: Oxford University Press; Sep 2015.
46. Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. Scand J Public Health 2011 Jul;39(7 Suppl):38-41. [doi: 10.1177/1403494810394717] [Medline: 21775349]
47. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis F, Tan SS, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013 Jun 15;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
48. Kirisits A, Redekop WK. The economic evaluation of medical devices: challenges ahead. Appl Health Econ Health Policy 2013 Feb;11(1):15-26. [doi: 10.1007/s40258-012-0006-9] [Medline: 23329383]
49. Drummond M, Tarricone R, Torbica A. Economic evaluation of medical devices. Oxford Res Encycl Econ Financ 2018 Mar 28:A. [doi: 10.1093/acrefore/9780190625979.013.105]
50. Statistikbanken - LONS20: Løn efter arbejdsfunktion, sektor, aflønningsform, lønmodtagergruppe, lønkomponenter og køn [Statistikbanken - LONS20: Wages estimated based on job function, sector, payment form, employment group, wage component and gender]. Statistics Denmark. 2018. URL: https://www.statistikbanken.dk/statbank5a/default.asp?w=1536 [accessed 2020-06-10]

51. Levetider [Service lives]. Agency for Modernisation - Ministry of Finance. URL: https://oes.dk/oekonomi/oeav/regnskabsregler/generelle-bogfoeringsbestemmelser/levetider/ [accessed 2020-06-18]
52. Den samfundsøkonomiske diskonteringsrente [The economic discount rate]. Ministry of Finance. URL: https://fm.dk/media/10174/dokumentationsnotat-_-den-samfundsoekonomiske-diskonteringsrente.pdf [accessed 2021-09-17]
53. Ehlers LH, Vestergaard AS. Costing in Health Economic Evaluation. Aalborg: Aalborg University Press; 2019:1-88.
54. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014 Dec;32(12):1157-1170 [FREE Full text] [doi: 10.1007/s40273-014-0193-3] [Medline: 25069632]
55. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Br Med J 2009 Jun 29;338:b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
56. Kirkwood B, Sterne J. Essential Medical Statistics. Malden, Mass: Wiley-Blackwell; 2003.
57. Franklin M, Enrique A, Palacios J, Richards D. Psychometric assessment of EQ-5D-5L and ReQoL measures in patients with anxiety and depression: construct validity and responsiveness. Qual Life Res 2021 Sep;30(9):2633-2647 [FREE Full text] [doi: 10.1007/s11136-021-02833-1] [Medline: 33835414]
58. van Asselt AD, Dirksen C, Arntz A, Giesen-Bloo J, Severens J. The EQ-5D: a useful quality of life measure in borderline personality disorder? Eur Psychiatry 2009 Mar;24(2):79-85. [doi: 10.1016/j.eurpsy.2008.11.001] [Medline: 19095421]
59. Bergmo TS. Using QALYs in telehealth evaluations: a systematic review of methodology and transparency. BMC Health Serv Res 2014 Aug 03;14:332 [FREE Full text] [doi: 10.1186/1472-6963-14-332] [Medline: 25086443]
60. Lauritsen K, Innocenti AD, Hendel L, Praest J, Lytje MF, Clemmensen-Rotne K, et al. Symptom recording in a randomised clinical trial: paper diaries vs. electronic or telephone data capture. Control Clin Trials 2004 Dec;25(6):585-597. [doi: 10.1016/j.cct.2004.09.001] [Medline: 15588745]
61. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. JMIR Mhealth Uhealth 2019 Oct 15;7(10):e13479 [FREE Full text] [doi: 10.2196/13479] [Medline: 31617851]
62. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
63. Abu-El-Noor N, Aljeesh Y, Bottcher B, Abu-El-Noor M. Impact of a mobile phone app on adherence to treatment regimens among hypertensive patients: a randomised clinical trial study. Eur J Cardiovasc Nurs 2021 Jun 29;20(5):428-435. [doi: 10.1177/1474515120938235] [Medline: 32631080]
64. Austin SF, Jansen JE, Petersen CJ, Jensen R, Simonsen E. Mobile app integration into dialectical behavior therapy for persons with borderline personality disorder: qualitative and quantitative study. JMIR Ment Health 2020 Jun 11;7(6):e14913 [FREE Full text] [doi: 10.2196/14913] [Medline: 32525488]
65. Innovation Fund Denmark. URL: https://innovationsfonden.dk/en [accessed 2021-10-19]

Abbreviations

BPD: borderline personality disorder
DBT: dialectical behavior therapy
EQ-5D-5L: EuroQol 5-Dimensions 5-Levels
QALY: quality-adjusted life year

Edited by R Kukafka; submitted 17.03.21; peer-reviewed by I Rodriguez, S Landes; comments to author 23.04.21; revised version received 17.06.21; accepted 27.07.21; published 11.11.21.

Please cite as:
Laursen SL, Helweg-Jørgensen S, Langergaard A, Søndergaard J, Sørensen SS, Mathiasen K, Lichtenstein MB, Ehlers LH
Mobile Diary App Versus Paper-Based Diary Cards for Patients With Borderline Personality Disorder: Economic Evaluation
J Med Internet Res 2021;23(11):e28874
URL: https://www.jmir.org/2021/11/e28874
doi:10.2196/28874
PMID:34762057

Original Paper

The Sociological Perspective of Users' Invisible Work: A Qualitative Research Framework for Studying Digital Health Innovations Integration

Dilara Vanessa Trupia[1,2], PhD; Alexandre Mathieu-Fritz[1], Prof Dr; Tu Anh Duong[2,3], MD, PhD

[1]LATTS, Univ Gustave Eiffel, CNRS, Ecole des Ponts, Marne-la-Vallée, France

[2]INSERM, Chaire Avenir Santé Numérique IMRB U955, Équipe 8, University of Paris-Est Créteil, Créteil, France

[3]AP-HP, Department of Dermatology, Hôpital Henri-Mondor, Créteil, France

Corresponding Author:
Dilara Vanessa Trupia, PhD
LATTS, Univ Gustave Eiffel, CNRS, Ecole des Ponts
5 boulevard Descartes, Cité Descartes
Bâtiment Bois de l'Etang
Marne-la-Vallée, F-77454
France
Phone: 33 618635423
Email: dilaratrupia@gmail.com

Abstract

Background: When new technology is integrated into a care pathway, it faces resistance due to the changes it introduces into the existing context. To understand the success or failure of digital health innovations, it is necessary to pay attention to the adjustments that users must perform to make them work, by reshaping the context and sometimes by altering the ways in which they perform activities. This adaptation work, most of which remains invisible, constitutes an important factor in the success of innovations and the ways in which they transform care practices.

Objective: This work aims to present a sociological framework for studying new health technology uses through a qualitative analysis of the different types of tasks and activities that users, both health professionals and patients, must perform to integrate these technologies and make them work in their daily routine.

Methods: This paper uses a three-part method to structure a theoretical model to study *users' invisible work*. The first part of the method includes a thematic literature review, previously published by one of the coauthors, of major sociological studies conducted on digital health innovations integration into existing care organizations and practices. The second part extends this review to introduce definitions and applications of the *users' invisible work* concept. The third part consists of producing a theoretical framework to study the concept according to the different contexts and practices of the users.

Results: The paper proposes four dimensions (organizational, interactional, practical, and experiential), each composed of a set of criteria that allow a comparative analysis of different users' work according to different health technologies.

Conclusions: This framework can be applied both as an analytical tool in a research protocol and as an agenda to identify less visible adoption criteria for digital health technologies.

(J Med Internet Res 2021;23(11):e25159) doi:10.2196/25159

KEYWORDS

digital health innovations; qualitative analysis; sociological framework; invisible work; patient work; user work; participatory health care; chronic illness; self-quantification

Introduction

Background

A large number of digital health technologies are developed every day, each being increasingly less expensive, faster, more connected, and *smarter* than the last. On the one hand, the use of telemedicine among health professionals comes with the promise of revolutionizing the organization of health care, which can now be accessed increasingly at a distance. On the other hand, data-driven applications and new quantification practices seem to foster a belief in greater autonomy for patients.

Increasingly equipped, connected, and measured patients are the active figures of an empowered generation that is becoming digitally engaged in their health by collecting, sorting, interpreting, and deleting many types of health and environment–related data [1]. Although these two examples seem to refer to very different realities, they illustrate a common fact that goes beyond their provisional and circumstantial distinctions: the supply of new digital technologies is only growing in the field of health, whereas it is challenging health care practices. For the sake of this argument, we will use the term *digital health innovations* to refer to the adoption of both professional-centered telemedicine or telehealth technologies (eg, teleconsultation, tele-expertise, telemonitoring) and patient-centered telecare technologies (eg, eHealth, mobile health, u-health, self-monitoring, self-help).

However, providing interesting technologies does not necessarily lead to their adoption or normalization. Some are successfully integrated into everyday practices, while others are used only by the circle of the first experimenters. What makes some of these innovations work? Which factors explain their success or failure? These questions are at the core of a very large body of multidisciplinary literature dealing with the issues of integrating new technologies into existing organizations and care practices [2]. This literature provides a whole set of criteria for assessing the chances of a successful integration [3]. Although it promotes macroscale analysis, it can sometimes overlook the actual practices and their concrete realities as presented to users. The purpose of this paper is to formulate a theoretical framework to better account for practice-related criteria in explaining the success of new health technologies. First, we should present and contextualize the work that we are mobilizing within the existing literature. Earlier studies conducted on telemedicine in the 1990s point mainly to the technical difficulties in explaining its slow diffusion. This focus suggests that it is sufficient to resolve these difficulties for telemedicine to be widely accepted [4]. Numerous studies conducted from the point of view of their acceptability or usefulness develop the idea that practices could be transposed as they are within the new framework. In this regard, May et al [5] state that “the existing literature on telemedicine [in 2001] has for the most part taken as its primary focus the utility and efficacy of the technology itself, as it is applied to particular clinical settings and problems. This is primarily clinical literature that is about establishing the safe practice of medicine using a diverse set of communications technologies.” This clinical approach, which seeks to evaluate telemedicine devices according to the principles of *evidence-based medicine* [6,7], is further coupled by numerous economic analysis models built around their cost-efficiency. However, it has since become clear that the development of innovations depends on much more varied and, more importantly, socially defined factors than merely the technical, clinical, or economic efficiencies. As such, these one-dimensional approaches tend to underestimate the importance of the institutional, organizational, and professional contexts in which these technologies are concretely integrated.

In this regard, many recent studies have focused on socio-organizational factors to assess the likelihood that a new technology will be integrated successfully into existing health care organizations and practices. Some works propose combining different socio-organizational aspects within a multidimensional assessment model [8]. For example, this is the case of the *Model for Assessment of Telemedicine*, which has been established in this field [9]. Although such models do indeed consider multiple aspects in the evaluation of digital health innovations, they are still far from capturing the plurality and complexity of the activities that need to be performed to make these technologies work concretely, on a daily basis, for very different actors who pursue highly varied objectives, in widely differing sociotechnical configurations. In fact, as shown by Pols [10] and, more broadly, by a large corpus of work in *Science and Technology Studies*, it is difficult to accurately predict the success or failure of innovations in telemedicine, nor the concrete integration modalities of digital technologies in existing health care practices. Users do not necessarily turn to these, out of a taste for technology (ie, technophilia). Moreover, a majority integrate them in a context full of uncertainty, where it is unclear how technologies will transform their practice. The only thing that can be predicted in this regard is the unpredictability of the adoption modes of these technologies. This unpredictability is based on the diversity of user profiles (eg, medical doctors, nurses, patients, helpers) and the types of technologies (eg, professional tools, connected objects, web platforms, and mobile apps). It is also based on various places (eg, hospitals or private practices, in cities or in small towns), sociotechnical environments (eg, resources, tools, and information systems), medical situations (eg, routine care, chronic disease monitoring, or emergency), and broader health care contexts (ie, during a health crisis) in which they are integrated, accessed, and used [11].

To understand the success of digital health innovations, it is necessary to examine the efforts or, more precisely, to quote Nicolini [12], “the work to make it work” provided by different stakeholders, not only by innovators and promoters but also by users, both health care professionals and patients. We suggest that all this *user work*, most of which remains invisible, is an important factor in explaining not only the phenomena of adoption of innovations but also the way in which they transform existing organizations and care practices. Here, we understand the notion of innovation as a continuous and evolutionary process in which users (and their practices) play a key role in the successful integration of new technologies.

Scope and Purpose

Sociological studies have shown that the adoption of new health technologies by users is neither straightforward nor given [13]. As soon as a new technology is integrated into a care pathway, it inevitably faces resistance and creates obstacles due to the wide range of changes introduced in the existing context. These changes can concern the social and spatial organization of health care, the division of medical and paramedical work, the interactions between their various actors, the work activities themselves, as well as the knowledge levels and professional identities. To understand the success or failure of digital health innovations, it is necessary to pay attention to this prior context, which is bound to evolve with the integration of new technologies. More precisely, it is necessary to evaluate the adjustments that users must do to make it work by adapting the

pre-existing context and sometimes by altering or even inventing new ways of doing things [14]. These adjustments represent a relatively constrained form of activity. However, without these, new technologies are difficult to normalize or can be abandoned.

A large body of work conducted in *Science and Technology Studies* provides valuable insights on the importance of these activities. This is done specifically by the Actor-Network Theory, which brings to light the unsuspected work of *translation* that must indeed be performed by innovators, both industrials who design and distribute new technologies and coordinators who seek to integrate them into their organizations [15-17]. Beyond traditional project management work, promoters must be able to *align* heterogeneous elements into a coherent whole. They must synchronize not only sociotechnical environments [18] but also practices and interactions that new technologies may suggest organizing in different ways. This alignment refers to work that is often invisible [19]. It requires building meaning (ie, *sensemaking*) and trust around these devices [20] and negotiating collective understanding among various categories of actors, including physicians, nurses, care assistants, patients, and family caregivers. These actors have *a priori* different interests that must be considered and converge to a minimum degree in order to support the normalization of new technologies. This is an eminently delicate task, not only because it requires articulating a variety of interpretations and modalities of action but also because new technologies affect autonomy, which is traditionally very important both for health professionals [21] and for their patients [22]; thus, it is difficult to negotiate.

These studies draw attention to the work that innovators and coordinators need to do to address the organizational complexity of integrating telemedicine devices in existing health care organizations. They also point out the *real work* needed to be done for these technologies to work on a daily basis, not only by promoters but also by users [19,23]. To make digital health technologies work, they must perform a series of additional activities that can be studied as a specific form of work [23]. Thus, to integrate new technologies into existing health care organizations, it is necessary to be able to measure *a priori* what is being asked of users [24]. In this regard, many approaches attempt to assess the acceptance of new technologies through functional, cognitive, or ergonomic analyses of their uses. However, on their own, these approaches are also struggling to produce a conceptual model whose ability to reflect the reality of practices depends on its interdisciplinary nature. A model that allows to do just that is a specific branch of sociology known as *practice studies* (or *practice-based studies*), which has resulted from the dialogue between multiple approaches, including not only *Science and Technology Studies* and innovation studies but also ethnomethodology and theories of distributed cognition that include more contributions of cognitive sciences and ergonomics [25]. Despite their differences, these disciplinary approaches all call for examination of how practices are engaged in and considered by the users themselves in their localized contexts. Inspired by these practice-based studies, this paper explains the success of digital health innovations through an analysis of the concrete activities that users, both health professionals and patients, must perform to make them work in a daily routine in their various socio-organizational contexts. Its purpose is to introduce the concept of *users' invisible work* and to develop a sociological framework for studying it in the case of very different technologies and across several dimensions.

Methods

This paper uses a three-part method to develop a sociological framework for studying *users' invisible work*. First, it includes a thematic literature review that was previously performed and published by one of the coauthors. In this review, Alexandre Mathieu-Fritz organized and discussed major themes emerging from the existing literature conducted by both French and Anglo-Saxon social scientists on digital health innovations integration into existing care organizations and practices [2].

The second part of the method is to extend this review to introduce definitions and applications of *users' invisible work* concept, which is coined at the crossroads of three conceptual models, each of which offers one keyword and one original work that serves as a starting point for building a framework. The first one is the *invisible work* model [26,27] used in the sociology of innovation, *Science and Technology Studies*, and sociology of work, to shed light on all efforts that players must deploy for an innovation to be successfully integrated in the existing organizations. The second model is that of *patient work* [23] coined in the sociology of professions and medical work, to recognize the key role of patients in their own health care. The third model is developed in a French corpus of sociology of work and economic sociology [28] and discussed in internet studies [29]. It refers to the unrecognized activities of consumers and web users that create value for private actors, both manufacturers and digital platforms. The purpose of this extension is not so much to present all these existing works as to identify the original contributions of these respective conceptual models in the study of different kind of work activities that users have to do to successfully integrate new technologies into their routines.

The third and last part of the method consists of producing a theoretical framework to study variations in the users' work according to different types of digital health technologies and users. This framework is built using empirical work performed by the co-authors themselves in three different fields, each of which refers to a particular configuration of digital health practices. It includes teleconsultations between health care professionals and patients (in dermatology, geriatrics, and mental health), tele-expertise between health care professionals (in dermatology), and self-monitoring by patients themselves (monitoring of diabetes and cystic fibrosis). The framework was thus constructed by putting different practice-based criteria identified in the first part of the method, in the thematic literature review, through the filter of these three very different cases. This paper does not present the results of these studies, which will be published separately. However, it relies on this empirical work to generate an analytical grid that can be applied to study work activities in very different technological configurations.

Results

Overview

Users' invisible work concept emanates directly from the perspective of the interactionist sociologist Anselm Strauss who coined the term *patient work* in the early 1980s. We first present the *patient work* model and its different applications in recent works on the use of digital health technologies by patients. We then extend this original definition not only to all technology users, both patients and professionals, but also to all kinds of work activities that remain invisible, including those related to more recent data-driven tracking apps. On the basis of this extended framework, we present an analytical grid that can be used both as an analytical tool in a research protocol and as a research agenda to assess the successful integration of digital health innovations into the existing health care organizations. The research protocol provides a tool to analyze this work through four dimensions related to the integration of new technologies (ie, organization of care, interaction between professionals as well as professionals and patients, clinical practice, and subjective experience of these users). As an agenda, it proposes to study the *users' invisible work* in very different technical configurations and socio-organizational contexts in which it is performed. This qualitative research package is expected to reveal different types of work activities as more subtle criteria to explain variations in uses according to users' specific contexts and to produce a comparative analysis between different type of work that various digital health technologies need them to do.

The Patient Work Model: Old Concepts, New Realities

In his work on medical worlds, Strauss drew attention to the fact that health care activities cannot be performed without the active participation of patients who have to perform a series of practical and cooperative tasks on a daily basis outside of health care facilities [23,30,31]. At first glance, it may seem peculiar to describe these activities as work. Generally, they are not considered as such, either by patients or health professionals [32]. However, this is justified in the case of chronic illnesses, which introduce profound changes in patients' daily practices and experiences [33]. Furthermore, these activities can be expected from chronic patients, and their key roles can be even recognized by professionals. According to Strauss, patients are central actors in the division of medical work [34]. They have to develop certain corporal attention and organize the daily management of their care (ie, *illness work*). Beyond concrete activities, they also have to develop different forms of reflexivity to reconfigure their inner experience of the illness [35], meaning the way they view and build their future lives (ie, *biographical work*).

In traditional care, *technical* tasks, which cover the care practices undertaken for the direct purpose of altering the course of the pathology, are separated from *other forms of work*. Among these, Strauss identifies *clinical safety work* (ie, anticipation, verification, evaluation, and, where necessary, correction), *machine work* (ie, maintenance, monitoring, use and, where necessary, repair), *comfort work* (ie, aiming to reduce pain or discomfort), *information work* (ie, requests, reports, and reassurances), and *sentimental work* (ie, improving emotions of patients and coping with the psychological effects) [31]. These different forms of work are profoundly affected by the integration of new technologies that displace the concrete efforts as well as bodily sensations and thoughts that underlie the management of the illness [36,37]. This implies new forms of task delegation where patients take charge of their own medical surveillance and become true *diagnostic agents* [38]. In the case of telemedicine, Oudshoorn observes, for example, a certain disciplinarization of patients who often end up establishing *self-management techniques* [19]. Hence, they must perform a certain number of diagnostic procedures, acquire new skills related to the use of devices and the interpretation of symptoms, and, *in conclusion*, translate this learning into practical knowledge that they can use in the daily care of their disease [39].

These observations are supported by recent research on data-driven self-tracking apps. For example, the work of Mathieu-Fritz and Guillot conducted on the case of diabetes and self-monitoring devices [40]. The authors illustrated how the use of this device introduces new forms of work, reflexivity, and self-knowledge associated with the illness experience. Through a comparative analysis of old and new glucose meters, the authors showed that the permanent sensor placed on the arm alleviated some of the usual constraints posed by the fingerstick capillary testing methods [40]. Each of these devices refers to a set of constraints that organize the illness experience and its daily management differently. These constraints are not only physical (ie, pain) and material (ie, maintenance and transport of equipment) but also organizational (ie, anticipation and planning), spatial (ie, conditions for use of the devices), symbolic (ie, disclosure of the illness), and social (ie, strategies of discretion and breaking of interactions). This comparison of different types of patient work reveals more subtle criteria for understanding variations in the use of continuous glucose readers: more discreet and rapid, they can be used in social spaces where glycaemia measurement was not previously practiced, allows for more frequent measurements to be taken, and develop a different approach to anticipating treatment. The different forms of activities that the devices allow to perform or avoid transform the reflexive work of patients and, thus, the way in which they administer their own care. These observations are supported by several studies conducted more broadly in the field of prevention and well-being, where new self-quantification practices are being developed [41]. The confrontation of individuals with a *quantified self* not only provides new information to guide individual choices in the management of care but also produces new representations that profoundly affect the subjective experience of the body, the illness, and its daily management [42].

The application of this old concept of patient work to new self-monitoring and self-care practices points to the reflexive nature of new connected devices that are used more frequently in the medical field. The very nature of this work is changing in accordance with the new principles of visibility and recognition of these activities [43]. That said, one must wonder to what extent the action of scanning a sensor, or, even more unconsciously, the action of leaving mobile apps programmed

by default, such as Apple's health watch, to collect and record data, can be studied as new forms of work. This debate, which is at the center of current concerns on the development of digital platforms, renews the interest and heuristic value of this concept to understanding how new *reflexive technologies* [44] affect the organization, practice, and experience of care [45,46]. From a critical perspective, some denounce the tendency of *techno-utopic* discourses to obscure the social, cultural, political, and economic dimensions of self-care technologies. Far from being neutral, these are, in fact, caught up in power relations [1,47]. This idea has been extensively developed in research inspired by *Science and Technology Studies*, which reports on the use of self-tracking devices as a reconfiguration of the physician-patient relationship through a new set of activities to generate and interpret data [48]. Further studies on daily data transmission devices to health care professionals show, for example, that patients can negotiate to redefine the objectives associated with the device by developing unexpected uses to prevent data transmission [49]. The merit of practice-based studies reveals itself through these works that study self-quantification as a particular form of work that must be performed *on* and *with* data [50], not only to contextualize them but also to articulate them with other forms of knowledge that supports the daily care practices [51]. This may represent one of the reasons why it is necessary to extend the analytical framework not only to various users (including health professionals) but also to different dimensions of their work, to include their reflexivity.

An Extended Framework for Studying Users' Invisible Work

Patient work is a powerful concept for highlighting and recognizing patients' role and engagement in their own care. However, an extension is necessary to use it in the assessment of digital health innovations that requires considering all users, not only patients but also health care professionals who must make these devices work during their daily practice. For instance, Oudshoorn shows through the case of telemonitoring that physician not only interpret electrocardiograms but also take on a series of tasks unforeseen by the designers. They assist and reassure patients in the use of medical technology, ease concerns, and contribute to building trust in medical technology [19,52,53]. For the author, this is akin to *affective work*, which is equivalent to the *sentimental work* in Strauss' classification [54]. Here, the notion of user work is clearly different from that of usability. The study of this work is not so much about the capacity or efficiency of uses as about the set of adjustments that users must make to resolve different types of conflicts that these technologies can introduce in the existing context. In that sense, studying users' work is first, to understand how the integration of new technology transforms existing work activities.

In that regard, research have shown that new digital health technologies do not simply equip practices from *outside* but contribute to transform them in a profound way, from the *inside*. Nicolini's work is particularly instructive in this regard. By *zooming in* on the practices and *zooming out* on the broader organizational context [55], the author highlights how work is redistributed among practitioners, as well as with artifacts (ie, *nonhumans*). In this context, the adoption of these devices depends on a series of learnings, settings, and adjustments that must be made in users' working practices [13]. A typical case is that of teleconsultations during which the roles played by each participant will be different from those usually played in a face-to-face setting [56]. Similar to other scholars, Nicolini observed that professional practices were transformed in situations where new forms of task distribution and delegation occur not only between physicians but also between physicians and paramedical professions (ie, nurses and caregivers) [57]. In some cases, important tasks that are considered to be central in the professional practice (ie, *true work* [58], as opposed to *dirty work* [59], which is considered to be peripheral) can be delegated to patients themselves.

These new forms of cooperation can also be observed among actors from different specialties who are placed at different hierarchical levels. In these configurations, they can contribute to the sharing of medical and clinical knowledge and enable certain delegates to *increase in expertise.* Moreover, it can also redefine the boundaries of their professional territory and identity [60]. Research on telemedicine devices show that "the question is not so much what the new activities allow, but rather to what extent they allow existing and appropriate forms of professional knowledge and practice to be put in place" [5]. However, this observation must be nuanced. Professionals may also get round the specific constraints of telemedicine (eg, physical distance and deprivation of *sensory inputs* [19,61]), for example, to develop new therapeutic techniques in the context of teleconsultations in mental health, by testing sooner than usual their *clinical intimacy* or by asking questions more frequently [62]. Thus, the integration of new technologies may end up producing new practices. Here, the technology itself becomes a full actor with whom one, not only redistributes existing work but also jointly produces new information that needs to be interpreted. These new forms of work become just as apparent to patients, for example, in the case of quantification devices that introduce new indicators in self-monitoring practices.

The patient work model must be extended to include not only existing work activities that have to evolve with the integration of new technologies but also new activities that users must perform and skills they need to acquire to use these technologies during their daily activities. Thus, the study of users' work invites a more systematic look at the articulation and coordination of different types of tasks within the socio-organizational context in which technologies are being integrated and used. The work of Mathieu-Fritz et al sheds light in this respect. First, the authors show that coordinating physicians play a key role in solving difficulties and problems that the designers did not foresee [52]. The authors also highlight different types of *framing* work with the purpose of establishing rules and guidelines and ensuring that users become autonomous [63-65]. Technical framing aims to teach how to use devices effectively (eg, synchronized use of mobile camera, dermatoscope, and spirometer), whereas social framing is about establishing ways to interact efficiently during teleconsultations (eg, presenting oneself precisely and systematically, speaking louder, and articulating better). Clinical framing defines the

protocols to follow (eg, patients' clinical history, auscultation, diagnosis, prescriptions, or indications), whereas organizational framing involves cultivating appropriate attitudes and behaviors (eg, being punctual, completing forms and, reminding appointments), to ensure that all actors are present at a given time on both sides of the camera.

The authors are particularly attentive to the additional coordination activities that medical and paramedical actors must perform for the day-to-day operation of these devices [27]. Referring to the work of Strauss, they emphasized the importance of *articulation work* [66,67]. In their research on the medical world, Strauss et al [68] define the social organization of care as a *negotiated order* that results from the constant efforts of the actors to produce, often informally, an agreement on the best ways to organize tasks daily. As Star [26] reminds us, this articulation work is beyond the scope of rational work organizations. It consists of organizing, coordinating, and combining different types of activities conducted by various actors (eg, a clinical examination, an x-ray, a blood or blood pressure test, prescription, and administration of medication) to ensure that "the staff's collective efforts add up to more than discrete and conflicting bits of accomplished work" [31]. The integration of new technologies has transformed coordination work. New coordination tasks are also required to ensure day-to-day operation. These activities are important, even sometimes decisive, so much so that certain works identify their intensity as a rejection factor of new devices [69].

Studying all this background work, most of which remains invisible, is crucial to account for the efforts required to make technologies work. However, most of these work forms are difficult to account for in the design and integration process of new technologies [70,71]. On the basis of the lessons learned from this literature, we propose to define the scope of the *users' invisible work* concept as all the concrete and reflexive activities that both health care professionals and patients perform to make digital health technologies work within their daily routines around health care. One of the original contributions of this framework is to emphasize the importance of reflexivity and experience forms that develop through local and subjective confrontations with the technologies in the trajectory of these innovations. The notion of reflexivity is used here in its two registers. Retrospective reflexivity refers to the ability to return from experiences and past events. Similar to immediate reflexivity, lessons learned from previous experiences serve to restructure action as they unfold [72]. The reflexive dimension is becoming increasingly important with data-driven apps, either because this work is increasingly automated by connected and intelligent objects, to the point that this is done sometimes entirely out of awareness or, on the contrary, because the data are increasingly visible to the patients who produce, interpret, and communicate it. In any case, one of the main contributions of this extended framework concerns precisely the way it suggests analyzing the work activities as they are performed by users, but also, and more importantly, their reflexivity (ie, what they think and feel as they perform them). It is only then that some nuances can be explored to understand resistance to technology. This is the case for some health professionals who reject new technologies not because they resist change but because they believe that these do not allow them to do what they consider to be the core of their work, their *real work* [73]. In other words, rejection can express a transposability issue where users consider that they cannot work or care with the same quality they would usually expect to have. This is clearly a separate issue from a social resistance to change.

A Sociological Research Package to Assess Digital Health Innovations

This theoretical framework can be used both as an analytical tool in a research protocol and as a qualitative research agenda to assess the chances that digital health technologies have to be successfully integrated into existing health care organizations. In a research protocol, for any given technology, it proposes to produce an overview of different actors who constitute the sociotechnical network around their use. All user types (eg, physicians, patients, and family caregivers) and material supports (ie, technical objects, instructions, and protocols) must be fully included as human and nonhuman actors, which must be aligned to some extent for innovations to be successfully integrated into daily routines [17]. This ecological approach must also be applied to cover a variety of tasks and activities without which innovation is difficult to develop or abandoned. In this regard, all forms of work (eg, articulation work, patient work, and information work) and their various definitions (eg, well-done or acceptable work, desired or satisfactory work and prevented work [74]) must be considered. Overall, the goal is to capture the diversity of the actors performing a range of different tasks and activities, both practical and reflexive in nature, to integrate new technologies into existing health care organizations (Textbox 1).

Textbox 1. An ecological approach to the diversity of digital technologies, users, and work.

Types of technology

- Equipment sets (telepresence station, teleconsultation booth, telemedicine trolley, etc)
- Connected medical tools (spirometer, stethoscope, otoscope, ultrasound scanner, electrocardiogram, dermatoscope, etc)
- Mobile apps (crowdsourcing, quantification, gaming, self-help, etc)
- Web-based platforms and software services (networking, forums, videoconference, etc)
- Connected objects (sensors, readers, automatons, robotics, AI, vocal assistants, chatbots, carebots, wearables, etc)

Types of users involved

- Health professionals (specialized or general practitioners, nurses, etc)
- Allied health professionals (paramedics, therapists, assistants, auxiliary personnel, social workers, etc)
- Patients (chronic, nonchronic or acute, emergency, etc)
- Healthy individuals (eg, sportsmen and women)
- Family caregivers (parents, partners, friends, etc)

Types of tasks and work

- Translation work (eg, alignment and sensemaking)
- Coordination work (eg, articulation, framing)
- Patient work (eg, illness work, sentimental work, and information work)
- Users work (concrete efforts such as taking measurements, ticking boxes, or sharing data, but also different reflexivity forms)

To do so, this framework investigates the circumstances under which this work is performed through four dimensions related to the integration of new technologies (ie, organizational, interactional, practical, and experiential). The first dimension, organizational, concerns coordination modalities of health care. The term organization refers to the physical and material environment in which digital health technologies are integrated and whose level of saturation can be an important factor in users' rejection. It also refers to all the additional coordination activities that users must perform, such as articulation work that realigns affected *lines of work*, and framing work that defines rules to normalize uses. This organizational dimension relates to broader changes in the care pathway due to new technologies that define the scope and composition of work groups differently. This is, for example, the case of teleconsultation that allows health care professionals to be brought together with different specialties and sometimes actors who were not participating in traditional consultations, thus contributing to the formation of new micro–work collectives and to the evolution of patient care pathways [60,65].

The second dimension is that of interactions between various actors within health care organizations and with patients. It draws attention to the place of these actors in the whole system of interactions that is mobilized around patients' care [75]. It first includes new modalities in which these actors interact with each other through new technologies that tend to open up the singular colloquium between the physician and patient and change the relationship between the two by adding new actors, both from the medical side and the patients' side. In this regard, Oudshoorn refers to a major transformation in the geography of responsibilities [76] based on a new spatial and temporal distribution of activities related to care delivery. In this *techno-geography*, new proximities may develop between patients and health care professionals [77]. Oudshoorn insists on the way in which the *digital proximity* established through new telemedicine devices highlights "protocol-driven communication, daily surveillance, and self-care" [52] as full-fledged dimensions of the interactions that forge remote care relationships. As it is, this new type of proximity sets aside the psychosocial dimensions of traditional face-to-face care, which requires considering extramedical aspects. These transformations also occur at a more symbolic level. For example, some studies show that hierarchical representations or expert legitimacy can be at stake between professionals who use digital health technologies. Some can associate the use of certain technologies with a lack of professionalism. For example, some mental health professionals tend to hide their teleconsulting practices, which are not well regarded in their professional environment [62]. These symbolic effects also concern patients whose use of these technologies can vary according to the social environment in which they have to make them work. This is the case for diabetic patients who have to use a glucometer that not only discloses their illness but also disrupts social interactions in which they can be engaged [40].

The third dimension of user work is clinical practice, that is, activities directly associated with the diagnosis and medical treatment of patients. It includes new forms of distribution and delegation of tasks between physicians and allied professions [57] and with patients [19] that may enable sharing medical and technical knowledge and increasing expertise. Some users may also end up working twice as much to integrate these devices (ie, patient's double work) [40]. Furthermore, this redistribution of tasks transforms the broader topography and temporality of care activities, both for the professionals and patients. In this

regard, Nicolini showed that the daily exercise of telemedicine is characterized by a *stretching out* of sociomaterial arrangements in space-time and consequently an *expansion* of health care activities [57]. This means that these reconfigurations go far beyond the simple spatiotemporal redistribution of existing activities. They contribute to transforming the social spaces in which these activities are performed. Oudshoorn showed how the homes of older adults who are hospitalized at home with the help of telemonitoring devices becomes a hybrid place in which conflictual logics of care and aesthetics can coexist [76]. This new geography profoundly affects the object and content of these activities. It transforms the relationships between health care professionals by redistributing their work differently, sometimes with artifacts.

The fourth dimension is the subjective experience of these users, which refers to their feelings and representations that they develop within and through their practice. In fact, the integration of new technologies alters not only concrete work activities during the clinical examination but also professional judgments involved in the formulation of a diagnosis. For instance, building trust seems to be a core issue for health professionals who develop an *experimental attitude* [40] toward these devices. As mentioned above in the case of teleconsultations, professionals who engage in such practices with an unclear status question the extent to which previous practices can be transposed into the new framework. For example, they consider the problem of accountability for encountered difficulties. They also consider the costs of this transposition in economic (related to the equipment that needs to be acquired and their compatibility or interoperability with the existing ones), organizational (related to the time and resources needed for preparation), cognitive (related to learning and training needs), and social and symbolic (related to the image conveyed by the device) terms. This is not so different for patients whose subjective experience also plays a key role in how they manage their own health care. Indeed, the experience of illness encompasses the patient's work, but goes far beyond it. It refers to the whole inner experience of the disease, to all that is felt (eg, bodily sensations) and thought (eg, what one would like to do, or, on the contrary, to avoid) subjectively. Overall, the experience of the disease refers to the social definition of oneself (eg, when one tries to hide one's illness and when one talks about it or shows it in some way). New technologies also challenge these reflexive activities. An example is given by Van Hout in his work on a telemonitoring device that has been used with homecare patients who were minimizing their symptoms or even omitting to tell palliative care nurses [37]. However, the device that allowed patients to assess the severity of their various symptoms on a scale of 0 to 5 has not always worked, as patients have criticized the device for reminding them of symptoms they did not yet have or were at risk of having (ie, display effect). Therefore, *forgetting* the device, and thus the disease, can become an important factor in their adoption by patients [78].

This paper proposes to study *users' invisible work* through these four dimensions. Table 1 presents these practice-based criteria in a chart that can be applied as a tool in research protocols. However, there is no single recipe or exclusive way of applying this grid, which is rather a collection of criteria that can be combined as needed, depending on the technologies, contexts of use, and users being studied.

Table 1. An analytical grid to study users works through its four dimensions.

Users' work and criteria	Application examples
Organizational configurations	
Ecology of artifacts	Interoperability between information systems; unforeseen problem-solving; degree of saturation as a factor of rejection
Additional coordination work	Framing work (social, technical, clinical and organizational) to establish the rules of use; intensity of the articulation work as a rejection factor
Recomposition of the care pathway	Scope and composition of new work collectives; forms and conditions of preventative actions
Interactional settings	
System and modalities of interaction	Opening of the singular colloquium of patient-physician to new actors; digital proximity with the patients
Forms of cooperation	Delegation of tasks and sometimes even "real work" to other professionals and artifacts; patients as diagnostic agents
Symbolic effects	Professionals' legitimacy among colleagues; patients' illness in their social environment
Clinical practices	
Topography and temporality	Duplication of the therapeutic space; "expansion" of health care activities; new constraints that organize patients' reflexive work
Learning or cognitive aspects	Knowledge barriers; "increase in expertise"; double work of patients; disciplinarization of patients
Consideration of the psychosocial aspects	Reduced in digital proximity
Subjective experiences	
Commitment and trust in fuzzy practices	Experimental attitude; Transposability issues; economic, organizational cognitive, social and symbolic costs of transposition; Accountability problems
"Forgetting" about the disease or device	New information produced by the device (display effect); representations that affect the experience and the daily management of illnesses

Furthermore, this framework can also be applied as an agenda that calls for further research on the integration phase of various digital health technologies into work systems (eg, connected devices, data collecting and displaying apps, and more infrastructural telemedicine projects, such as teleconsultation cabins). The purpose of this agenda is to redefine some of the *a priori* distinctions (ie, telehealth vs digital health) through variations that can be observed at the level of the different work forms that require users to accomplish in very different configurations. This agenda should include studies of technologies that are used in different spaces, both very localized and spontaneous ones, such as teleconsultations, and mobile and ubiquitous ones, such as self-tracking devices. This agenda should also contain studies on apps in which data can be more or less important to explore the various effects of reflexive technologies on the self-knowledge of patients associated with their illness experience. It must cover devices used in different medical conditions, for example, for rare and common pathologies, to identify how technology adoption modes vary with the rarity of a disease. A preliminary list of such variables is presented in Table 2, which aims to cover a very wide range of configurations, including spatiotemporal, technical, social, and medical variations, to develop this comparative approach between different work forms required in the context of their uses.

Table 2. Variables in the use of digital health technologies.

Variables	Description
Spatial-temporal	
Topography	Single place (eg, fixed teleconsultation cabin), semi mobile (eg, telemedicine trolley), multiple place (eg, mobile apps)
Frequency and punctuality	Frequent (eg, multiple times a day) or rare (eg, once a month), precise (eg, on appointment) or approximate (eg, each morning)
Duration	Continuous (eg, wearable devices and sensors) or punctual (eg, teleconsultations)
Technical	
Autonomy	Automated (eg, parameterization) or triggered actions (eg, synchronization)
Visibility	Visible (eg, reader) or invisible (eg, implants)
Connectivity	Autonomous (eg, pedometer) or connected devices (eg, smart watches)
Artifactuality and agentivity	Cognitive (eg, stocking data) or interactive (eg, following feeds), passive (eg, visual representations) or active (eg, alerts and notifications)
Social	
Context of use	Routine check-up (eg, blood work), emergency (eg, allergy crises) or well-being (eg, sport diet)
Modality of use	Individual (eg, self-help, measurements, etc.) or collective (data sharing applications, dyadic or triadic tele-consultations, etc)
Type of actors	Professional caregivers (interaction with physicians, nurses, etc) or family members (assistance by parents, friends, etc)
Medical	
Pathology	Chronic (eg, diabetes) or acute (eg, heart spasm), rare (eg, cystic fibrosis) or common (eg, kidney disease)
Specialty	Primary care (eg, general medicine), secondary care (eg, specialists), tertiary care and hospitalization (eg, surgery)
Type of treatment	Preventive (eg, dietary programs), curative (eg, physical therapy), palliative or hospice (eg, cancer treatment)

With this agenda, it becomes possible to produce a comparative study of users' work variations in the case of different digital health practices and explain the success or failure of new technologies, according to more or less data-driven, real-time, frequent, or visible nature of activities they require users to do.

Discussion

Principal Findings

This paper draws attention to the plurality of most invisible and unrecognized tasks and activities that need to be performed by professionals and patients to make these devices work in various contexts. As a result, it proposes a theoretical framework that allows the assessment of digital health innovations by studying these work activities under different settings and the transformations they introduce through four main dimensions: organizations, practices, interactions, and experiences. For any given technology, it can be used as a tool in a research protocol to study concrete work activities performed by users to integrate them into their daily lives. It can also be applied as a research agenda that covers a wide range of technological configurations to develop a comparative approach through practice-based criteria.

Main Contributions

This theoretical framework makes three main contributions to the literature. First, it reports on professional practices and patient experiences jointly and in an articulated way. Second, it seeks to guide analytical practices by further operationalizing the theoretical approaches of practice-based studies in a methodological framework to help better understand the successful integration of new technologies into existing health organizations. Third, it redefines traditional categories such as *technology acceptance* or *resistance to change*, through an analysis that remains more faithful to the activity and the concrete reality of the users themselves.

Acknowledgments

This paper has been prepared as part of the research chair *Avenir Santé Numérique* launched by *Fondation de l'Avenir* and *University Paris-Est Créteil*, directed by TAD and financed by the *Avenir Santé Numérique* funds.

Conflicts of Interest

None declared.

References

1. Lupton D. Towards critical digital health studies: Reflections on two decades of research in health and the way forward. Health (London) 2016 Jan;20(1):49-61. [doi: 10.1177/1363459315611940] [Medline: 26487686]
2. Mathieu-Fritz A, Gaglio G. In search of the sociotechnical configurations of telemedicine. Réseaux 2018;1(207):27-63. [doi: 10.3917/res.207.0027]
3. May C, Harrison R, Finch T, MacFarlane A, Mair F, Wallace P. Understanding the normalization of telemedicine services through qualitative evaluation. J Am Med Inform Asso 2003;10(6):596-604 [FREE Full text] [doi: 10.1197/jamia.M1145] [Medline: 12925553]
4. Harrison RM, Clayton W, Wallace P. Is there a role for telemedicine in an urban environment? J Telemed Telecare 1997;3 Suppl 1:15-17. [doi: 10.1258/1357633971930832] [Medline: 9218369]
5. May C, Gask L, Atkinson T, Ellis N, Mair F, Esmail A. Resisting and promoting new technologies in clinical practice: The case of telepsychiatry. Soc Sci Med 2001 Jun;52(12):1889-1901. [doi: 10.1016/s0277-9536(00)00305-1] [Medline: 11352414]
6. Bashshur RL, Shannon GW, Smith BR, Woodward MA. The empirical evidence for the telemedicine intervention in diabetes management. Off J Am Telemed Asso 2015 May;21(5):321-354 [FREE Full text] [doi: 10.1089/tmj.2015.0029] [Medline: 25806910]
7. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Off J Am Telemed Asso 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
8. Williams T, May C, Mair F, Mort M, Gask L. Normative models of health technology assessment and the social production of evidence about telehealth care. Health Policy 2003 Apr;64(1):39-54. [doi: 10.1016/s0168-8510(02)00179-3] [Medline: 12644327]
9. Kidholm K, Clemensen J, Caffery LJ, Smith AC. The Model for Assessment of Telemedicine (MAST): A scoping review of empirical studies. J Telemed Telecare 2017 Oct;23(9):803-813. [doi: 10.1177/1357633X17721815] [Medline: 28758525]
10. Pols J. Care at a Distance: On the Closeness of Technology. Amsterdam: Amsterdam University Press; 2012.
11. 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]
12. Nicolini D. The work to make telemedicine work: A social and articulative view. Soc Sci Med 2006 Jun;62(11):2754-2767. [doi: 10.1016/j.socscimed.2005.11.001] [Medline: 16343724]
13. Akrich M, Méadel C. Problématiser la question des usages. Sciences Sociales et Santé. 2004. URL: https://www.persee.fr/doc/sosan_0294-0337_2004_num_22_1_1606 [accessed 2021-09-21]
14. Akrich M, Callon M, Latour B. Les utilisateurs, acteurs de l'innovation. In: Sociologie de la Traduction : Textes Fondateurs. Paris: Presses des Mines; 2013:253-265.
15. Callon M. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. The Sociological Review 2015 May 29;32(1_suppl):196-233. [doi: 10.1111/j.1467-954x.1984.tb00113.x]
16. Akrich M, Callon M, Latour B. Sociologie de la Traduction. Textes Fondateurs. Paris: Presses des Mines; 2006.
17. Akrich M, Callon M, Latour B. A quoi tient le succès des innovations ? 1 : L'art de l'intéressement; 2 : Le choix des porte-parole. Gérer et Comprendre Annales des Mines 1988:4-17-14-29. [doi: 10.3917/geco]
18. Sicotte C, Lehoux P. La télémédecine et la transformation des formes organisationnelles de l'hôpital. In: L'hôpital en Restructuration : Regards Croisés Sur la France et le Québec. Montréal: Les Presses universitaires de Montréal; 2005:163-184.
19. Oudshoorn N. Diagnosis at a distance: The invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol Health Illness 2008 Mar;30(2):272-288 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01032.x] [Medline: 18290936]
20. Barlow J, Bayer S, Curry R. Implementing complex innovations in fluid multi-stakeholder environments: Experiences of 'telecare'. Technovation 2006 Mar;26(3):396-406. [doi: 10.1016/j.technovation.2005.06.010]
21. Freidson E. Profession of Medicine: A Study of the Sociology of Applied Knowledge. Chicago: University of Chicago Press; 1988.
22. Lupton D. Beyond techno-utopia: Critical approaches to digital health technologies. Societies 2014 Dec 08;4(4):706-711. [doi: 10.3390/soc4040706]
23. Strauss AL, Fagerhaugh S, Suczek B, Wiener C. The work of hospitalized patients. Soc Sci Med 1982 Jan;16(9):977-986. [doi: 10.1016/0277-9536(82)90366-5]
24. Yin K, Jung J, Coiera E, Laranjo L, Blandford A, Khoja A, et al. Patient work and their contexts: Scoping review. J Med Internet Res 2020 Jun 02;22(6):e16656 [FREE Full text] [doi: 10.2196/16656] [Medline: 32484449]
25. Licoppe C. In the "activity square": International perspectives on labor, work and activity. Sociologie du Travail 2009 Nov;51:155-169. [doi: 10.1016/j.soctra.2009.06.024]
26. Star S. The sociology of the invisible: The primacy of work in the writings of Anselm Strauss. In: Maines D, editor. Social Organization and Social Process: Essays in Honor of Anselm Strauss. Hawthorne, NY: Aldine de Gruyter; 1991:265-283.

27. Star SL, Strauss A. Layers of silence, arenas of voice: The ecology of visible and invisible work. Comput Support Coop Work (CSCW) 1999 Mar;8(1-2):9-30. [doi: 10.1023/a:1008651105359]
28. Dujarier M. The three sociological types of consumer work. J Cons Culture 2014 Apr 08;16(2):555-571. [doi: 10.1177/1469540514528198]
29. Cardon D, Casilli A. Qu'est-ce que le Digital Labor ?. Bry-sur-Marne: INA Éditions; 2015.
30. Corbin J, Strauss A. Managing chronic illness at home: Three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
31. Strauss A. Social Organization of Medical Work. Chicago: University of Chicago Press; 1985.
32. Wiener C. Untrained, unpaid, and unacknowledged: The patient as worker. Arthritis Care Res 1989 Mar;2(1):16-21. [doi: 10.1002/anr.1790020105] [Medline: 2488117]
33. Bury M. Chronic illness as biographical disruption. Sociol Health Illness 1982 Jul;4(2):167-182 [FREE Full text] [Medline: 10260456]
34. Strauss A. La Trame de la Négociation: Sociologie Qualitative et Interactionnisme. Paris: Harmattan; 1992.
35. Pierret J. The illness experience: State of knowledge and perspectives for research. Sociol Health Illness 2003;25:4-22 [FREE Full text] [doi: 10.1111/1467-9566.t01-1-00337] [Medline: 14498927]
36. Strauss AL, Fagerhaugh S, Suczek B, Wiener C. Patients' work in the technologized hospital. Nurs Outlook 1981 Jul;29(7):404-412. [Medline: 6910640]
37. Hout AV, Willems D, Hettinga M, Pols J, Libbrecht E. Why expectations of telecare are often wrong. Réseaux 2018;1(207):95-121. [doi: 10.3917/res.207.0095]
38. Mort M, Finch T, May C. Making and Unmaking Telepatients: Identity and governance in new health technologies. Sci Technol Hum Val 2007 Dec 19;34(1):9-33. [doi: 10.1177/0162243907311274]
39. Pols J. Knowing patients: Turning patient knowledge into science. Sci Technol Hum Val 2013 Sep 13;39(1):73-97. [doi: 10.1177/0162243913504306]
40. Mathieu-Fritz A, Guillot C. Diabetes self-monitoring devices and transformations in "patient work". Revue d'anthropologie des connaissances 2017;11(4):641-675. [doi: 10.3917/rac.037.k]
41. Gilmore JN. Everywear: The quantified self and wearable fitness technologies. New Media Society 2016 Jul 10;18(11):2524-2539. [doi: 10.1177/1461444815588768]
42. Pantzar M, Ruckenstein M. Living the metrics: Self-tracking and situated objectivity. Digital Health 2017;3:1-10 [FREE Full text] [doi: 10.1177/2055207617712590] [Medline: 29942604]
43. Fiske A, Prainsack B, Buyx A. Data Work: Meaning-making in the era of data-rich medicine. J Med Internet Res 2019 Jul 09;21(7):e11672 [FREE Full text] [doi: 10.2196/11672] [Medline: 31290397]
44. Licoppe C. Formes de la présence et circulations de l'expérience. De Jean-Jacques Rousseau au « Quantified Self ». Réseaux 2013;182(6):21-55. [doi: 10.3917/res.182.0021]
45. Lupton D. Self-Tracking Modes: Reflexive self-monitoring and data practices. SSRN Electronic J 2014:e19. [doi: 10.2139/ssrn.2483549]
46. Cahour B, Licoppe C. Confrontations with traces of one's own activity. Revue d'anthropologie des connaissances 2010;4(2):243-253. [doi: 10.3917/rac.010.000a]
47. Lupton D. Critical perspectives on digital health technologies. Sociol Compass 2014 Dec 04;8(12):1344-1359. [doi: 10.1111/soc4.12226]
48. Bruni A, Rizzi C. Looking for Data in Diabetes Healthcare: Patient 2.0 and the re-engineering of clinical encounters. Sci Technol Stud 2013 Jan 01;26(2):29-43. [doi: 10.23987/sts.55297]
49. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: Negotiations in the ICT-mediated patient–provider relationship. Health Sociol Rev 2016 Aug 05;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]
50. Almalki M, Gray K, Martin-Sanchez F. Activity theory as a theoretical framework for health self-quantification: A systematic review of empirical studies. J Med Internet Res 2016 May 27;18(5):e131 [FREE Full text] [doi: 10.2196/jmir.5000] [Medline: 27234343]
51. Lupton D, Pink S, LaBond C, Sumartojo S. Digital traces in context : Personal data contexts, data sense and self-tracking cycling. Int J Commun 2018;12:647-665 [FREE Full text]
52. Oudshoorn N. Physical and digital proximity: Emerging ways of health care in face-to-face and telemonitoring of heart-failure patients. Sociol Health Illness 2009 Apr;31(3):390-405 [FREE Full text] [doi: 10.1111/j.1467-9566.2008.01141.x] [Medline: 19144084]
53. Oudshoorn N. Telecare Technologies and the Transformation of Healthcare. UK: Palgrave Macmillan; 2011.
54. Strauss A, Fagerhaugh S, Suczek B, Wiener C. Sentimental work in the technologized hospital. Sociol Health Illness 1982 Nov;4(3):78-254 [FREE Full text] [doi: 10.1111/1467-9566.ep10487954] [Medline: 10260459]
55. Nicolini D. Zooming In and Out: Studying practices by switching theoretical lenses and trailing connections. Organ Stud 2010 Jan 14;30(12):1391-1418. [doi: 10.1177/0170840609349875]
56. Pappas Y, Seale C. The opening phase of telemedicine consultations: An analysis of interaction. Soc Sci Med 2009 Apr;68(7):1229-1237. [doi: 10.1016/j.socscimed.2009.01.011] [Medline: 19201514]

57. Nicolini D. Stretching out and expanding work practices in time and space: The case of telemedicine. Hum Rel 2016 Apr 22;60(6):889-920. [doi: 10.1177/0018726707080080]
58. Bidet A. Qu'est-ce que le vrai boulot ? Le cas d'un groupe de techniciens. Sociétés Contemporaines 2010;2(78):115-135. [doi: 10.3917/soco.078.0115]
59. Hughes EC. Good people and dirty work. Soc Problem 1962 Jul;10(1):3-11. [doi: 10.1525/sp.1962.10.1.03a00010]
60. Mathieu-Fritz A, Esterle L. Changes in professional practices during medical teleconsultations. Revue Française de Sociologie 2013;54(2):303-329. [doi: 10.3917/rfs.542.0303]
61. Pols J. Wonderful Webcams: About active gazes and invisible technologies. Sci Technol Hum Val 2010 Aug 03;36(4):451-473. [doi: 10.1177/0162243910366134]
62. Mathieu-Fritz A. Teleconsultation in mental health. Réseaux 2018;1(207):123-164. [doi: 10.3917/res.207.0123]
63. Esterle L, Mathieu-Fritz A, Espinoza P. L'impact des consultations à distance sur les pratiques médicales. Vers un nouveau métier de médecin ? Revue Francaise Des Affaires Sociales 2011;1(2):79. [doi: 10.3917/rfas.112.0063]
64. Mathieu-Fritz A, Esterle L. Les transformations des pratiques professionnelles lors des téléconsultations médicales: Coopération interprofessionnelle et délégation des tâches. Revue Française de Sociologie 2013;54(2):303-329. [doi: 10.3917/rfs.542.0303]
65. Esterle L, Mathieu-Fritz A. Teleconsultation in geriatrics: Impact on professional practice. Int J Med Inform 2013 Aug;82(8):684-695. [doi: 10.1016/j.ijmedinf.2013.04.006]
66. Strauss A. The articulation of project work: An organizational process. Sociol Q 1988;29(2):163-178. [doi: 10.1111/j.1533-8525.1988.tb01249.x]
67. Corbin JM, Strauss AL. The articulation of work through interaction. Sociol Q 1993;34(1):71-83. [doi: 10.1111/j.1533-8525.1993.tb00131.x]
68. Strauss A. Negotiations: Varieties, Contexts, Processes and Social Order. San Francisco: Jossey-Bass Inc; 1978.
69. Sicotte C, Lehoux P. Teleconsultation: Rejected and emerging uses. Methods of Information in Medicine 2003;42(4):451-457. [Medline: 14534649]
70. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: Connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10. [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
71. Valdez RS, Holden RJ, Novak LL, Veinot TC. Technical infrastructure implications of the patient work framework. J Am Med Inform Assoc 2015 Apr;22(e1):213-215 [FREE Full text] [doi: 10.1093/jamia/ocu031] [Medline: 25665705]
72. Schön D. The Reflective Practitioner: How Professionals Think in Action. New York: Basic Books; 1983.
73. Bidet A. L'engagement dans le travail. Qu'est-ce que le vrai boulot ?. Paris: Presses Universitaires de France; Jun 15, 2012.
74. Clot Y. Le Travail à Cœur : Pour en Finir Avec Les Risques Psychosociaux. Paris: Découverte; 2010.
75. Hughes E. The Sociological Eye. New Brunswick, US: Routledge; 1984.
76. Oudshoorn N. How places matter: Telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012 Feb;42(1):121-142. [doi: 10.1177/0306312711431817] [Medline: 22530385]
77. Pols J, Willems D. Innovation and evaluation: Taming and unleashing telecare technology. Sociol Health Illness 2011;33(3):484-498 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01293.x] [Medline: 21241338]
78. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You Get Reminded You're a Sick Person": Personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]

Edited by R Kukafka, G Eysenbach; submitted 20.10.20; peer-reviewed by J Clemensen, S Myneni; comments to author 07.12.20; revised version received 22.01.21; accepted 31.07.21; published 04.11.21.

Please cite as:
Trupia DV, Mathieu-Fritz A, Duong TA
The Sociological Perspective of Users' Invisible Work: A Qualitative Research Framework for Studying Digital Health Innovations Integration
J Med Internet Res 2021;23(11):e25159
URL: https://www.jmir.org/2021/11/e25159
doi:10.2196/25159
PMID:34734832

Original Paper

Natural Language Processing and Machine Learning Methods to Characterize Unstructured Patient-Reported Outcomes: Validation Study

Zhaohua Lu[1*], PhD; Jin-ah Sim[2,3*], PhD; Jade X Wang[1], PhD; Christopher B Forrest[4], MD, PhD; Kevin R Krull[2], PhD; Deokumar Srivastava[1], PhD; Melissa M Hudson[5], MD; Leslie L Robison[2], PhD; Justin N Baker[5], MD; I-Chan Huang[2], PhD

[1]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, United States

[2]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, United States

[3]School of AI Convergence, Hallym University, Chuncheon, Republic of Korea

[4]Roberts Center for Pediatric Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[5]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, United States

[*]these authors contributed equally

Corresponding Author:
I-Chan Huang, PhD
Department of Epidemiology and Cancer Control
St. Jude Children's Research Hospital
MS 735, 262 Danny Thomas Pl
Memphis, TN, 38105
United States
Phone: 1 9015958369
Email: I-Chan.Huang@STJUDE.ORG

Abstract

Background: Assessing patient-reported outcomes (PROs) through interviews or conversations during clinical encounters provides insightful information about survivorship.

Objective: This study aims to test the validity of natural language processing (NLP) and machine learning (ML) algorithms in identifying different attributes of pain interference and fatigue symptoms experienced by child and adolescent survivors of cancer versus the judgment by PRO content experts as the gold standard to validate NLP/ML algorithms.

Methods: This cross-sectional study focused on child and adolescent survivors of cancer, aged 8 to 17 years, and caregivers, from whom 391 meaning units in the pain interference domain and 423 in the fatigue domain were generated for analyses. Data were collected from the After Completion of Therapy Clinic at St. Jude Children's Research Hospital. Experienced pain interference and fatigue symptoms were reported through in-depth interviews. After verbatim transcription, analyzable sentences (ie, meaning units) were semantically labeled by 2 content experts for each attribute (physical, cognitive, social, or unclassified). Two NLP/ML methods were used to extract and validate the semantic features: bidirectional encoder representations from transformers (BERT) and Word2vec plus one of the ML methods, the support vector machine or extreme gradient boosting. Receiver operating characteristic and precision-recall curves were used to evaluate the accuracy and validity of the NLP/ML methods.

Results: Compared with Word2vec/support vector machine and Word2vec/extreme gradient boosting, BERT demonstrated higher accuracy in both symptom domains, with 0.931 (95% CI 0.905-0.957) and 0.916 (95% CI 0.887-0.941) for problems with cognitive and social attributes on pain interference, respectively, and 0.929 (95% CI 0.903-0.953) and 0.917 (95% CI 0.891-0.943) for problems with cognitive and social attributes on fatigue, respectively. In addition, BERT yielded superior areas under the receiver operating characteristic curve for cognitive attributes on pain interference and fatigue domains (0.923, 95% CI 0.879-0.997; 0.948, 95% CI 0.922-0.979) and superior areas under the precision-recall curve for cognitive attributes on pain interference and fatigue domains (0.818, 95% CI 0.735-0.917; 0.855, 95% CI 0.791-0.930).

Conclusions: The BERT method performed better than the other methods. As an alternative to using standard PRO surveys, collecting unstructured PROs via interviews or conversations during clinical encounters and applying NLP/ML methods can facilitate PRO assessment in child and adolescent cancer survivors.

(J Med Internet Res 2021;23(11):e26777) doi:10.2196/26777

KEYWORDS

natural language processing; machine learning; PROs; pediatric oncology

Introduction

Pediatric Cancer and Patient-Reported Outcomes

Innovative anticancer therapies have significantly improved the 5-year survival rates of pediatric and adolescent patients with cancer in the United States [1-3]. However, toxic treatment often causes long-term sequelae (eg, physical and psychological morbidities and premature mortality [4-8]), which contribute to poor patient-reported outcomes (PROs) and impaired quality of life [8,9]. Poor PROs, such as fatigue, pain, psychological distress, and neurocognitive problems, are prevalent in survivors of cancer aged <18 years [10-12]. Approximately 50% of young survivors of childhood cancer experience severe fatigue [10,12,13] or pain [12,14], and both can worsen as survivors become older [15]. Assessing PROs from survivors and caregivers can complement clinical assessments, suggest potential adverse medical events, and facilitate the provision of appropriate interventions [16,17].

Unstructured PROs

Conventionally, PROs are collected from childhood survivors of cancer during follow-up care using standard surveys with prespecified content of PROs. Given busy clinic schedules, survivors may be unable or unwilling to complete surveys. Performing interviews or initiating conversations by clinicians are alternative methods of collecting PROs. However, PROs collected by this method are qualitative or unstructured in nature, which requires specific techniques for data processing and analysis. Natural language processing (NLP), a discipline of linguistics, information engineering, and artificial intelligence, initially designed for processing a large amount of natural language data, provides an innovative avenue for PRO research with potential clinical applications [18]. However, the validity of applying this method to evaluate PROs in oncology is understudied.

Application of NLP for PRO Analysis

NLP techniques have been applied to process unstructured or nonquantitative clinical data in medical notes for classifying or predicting health status (eg, risk of heart disease and stage of cancer) through information extraction, semantic representation learning, and outcome prediction [19]. Recently, NLP applications have been extended to unstructured PRO and symptom data stored in electronic medical records (EMRs) [20,21]. A review study [22] found that most previous NLP applications for unstructured PRO data largely focused on rule-based classifications (eg, extracting prespecified keywords or phrases from free text to identify cancer-related symptoms [23]), followed by machine learning (ML) approach (eg, conditional random field model [20], support vector machine [SVM] [24], and boosting regression tree [25]) to analyze associations with clinical outcomes.

The method of capturing the features of unstructured PROs is an emerging area of research [26]. Compared with rule-based extraction, the ML/deep learning–based NLP methods, including the context-independent or static (eg, term frequency–inverse document frequency [TF-IDF] [27], global vectors for word representation [GloVe] [28], and Word2vec [29]), and context-dependent or dynamic (eg, bidirectional encoder representations from transformers [BERT]; [30]) distributed representation methods are more suitable for processing unstructured PROs. Typically, context-dependent methods can capture the meaning of polysemous words, which substantially improves the flexibility and validity of analyzing unstructured PRO data.

Objective

To facilitate clinical decisions, our long-term goal is to collect PROs from survivor-caregiver-clinician conversations and apply NLP/ML methods to characterize meaningful PROs. Through in-depth interviews with childhood survivors of cancer and caregivers, this study evaluates the validity of using different novel NLP/ML methods (Word2vec/ML and BERT) to characterize 2 most common symptom domains (pain interference and fatigue) in child and adolescent survivors of cancer. The interview data were semantically labeled and coded by PRO content experts as the gold standard to represent specific symptom problems (defined as symptom attributes). In contrast to the static methods (ie, Word2vec/ML), we hypothesize that the use of dynamic methods (ie, BERT) would yield superior model performance.

Methods

Study Participants

Study participants were survivors of pediatric cancer and their caregivers recruited from the After Completion of Therapy Clinic at St. Jude Children's Research Hospital (*St Jude* hereafter) in Tennessee, United States, between August and December 2016. Eligible participants were identified from a list of survivors scheduled for annual follow-up and confirmed their eligibility through EMRs. We recruited survivors aged 8 to 17 years of age at annual follow-up, at least 2 years off therapy, and at least 5 years from initial cancer diagnosis. We excluded survivors who had acute or life-threatening conditions and required immediate medical care. We recruited caregivers who were the most knowledgeable of the survivor's health status and could speak or read English. Assent from survivors and consent from caregivers was obtained. The research protocol was approved by the institutional review board of St Jude.

In-Depth Interview and Data Abstraction

This investigation builds on our previous study that elucidated the contents of 5 PRO domains (pain interference, fatigue, psychological stress, stigma, and meaning and purpose) related to pediatric cancer from survivors and caregivers [15]. We

randomly assigned 2 domains to each survivor and 2 to 3 domains to each caregiver. PRO domains were assigned randomly to each survivor and caregiver to elucidate PRO contents from both survivors and caregivers rather than comparing PRO discordances between dyadic participants. Diagnostic and clinical information was abstracted from EMRs. We designed separate interview guides (Multimedia Appendices 1 and 2) with probes for each PRO domain, audio-recorded the interviews, transcribed interviews verbatim, and abstracted meaningful and interpretable sentences (ie, "meaning units") [15].

Expert-Labeled Outcomes as the Gold Standard

We used the methods developed in our previous studies to code the concepts of symptomatic problems collected from interviews and assigned the concepts to specific attributes [15,31]. Specifically, we began with abstracting the sentences or paragraphs collected from the interviews that are relevant to the experiences with particular symptomatic problems, such as presence, frequency, or intensity, and how these symptomatic problems affect daily activities (defined as meaning units) and then mapped the meaning units to analyzable, interpretable formats that represent the contents of items included in the Patient-Reported Outcomes Measurement Information System (PROMIS) banks [32] (defined as meaningful concepts). Subsequently, we labeled the meaningful concepts by distinct concepts, including physical, cognitive, and social (defined as attributes) concepts.

The associations among meaning units, meaningful concepts, and corresponding attributes are illustrated in Multimedia Appendix 3. For example, in the pain interference domain, when a survivor stated that "Can't play, and go outside when I have a headache," we mapped this meaning unit to the meaningful concept "Hard to do sports or exercise when had pain," and then labeled this meaningful concept as the *physical* attribute. For the fatigue domain, when a survivor stated that "It's hard to get my school work done when I'm tired," we mapped this meaning unit to the meaningful concept "Hard to keep up with schoolwork" and then labeled this meaningful concept as the *cognitive* attribute.

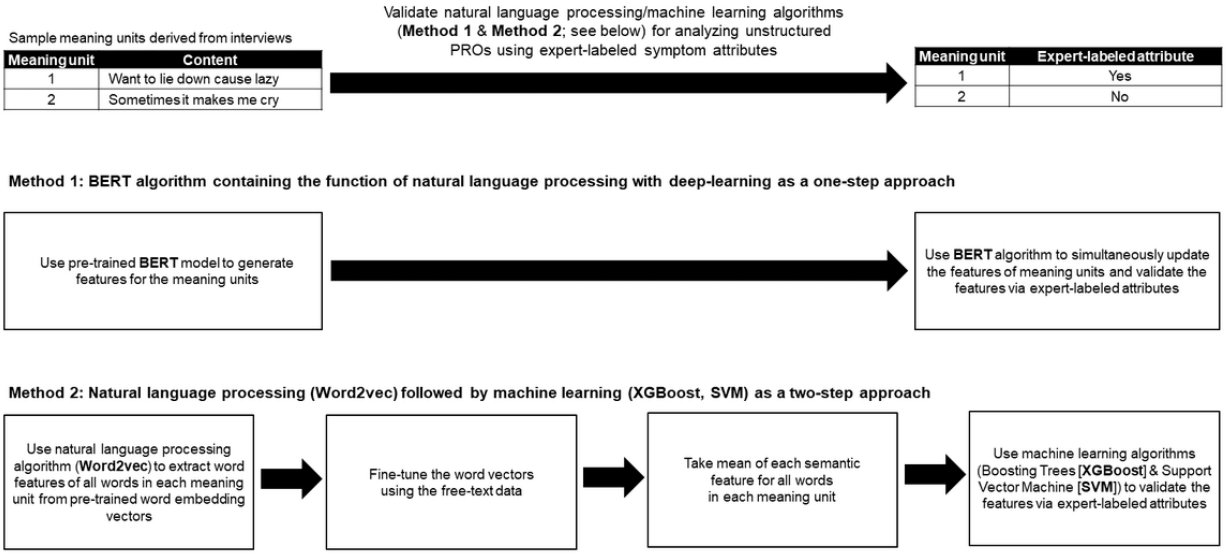
In addition, 2 PRO content experts (JLC and CMJ) independently reviewed the content of each meaning unit derived from the symptom domains and mapped each meaning unit to the content of individual items listed in the PROMIS pain interference and fatigue item banks [32]. In total, 391 and 423 meaning units representing pain interference and fatigue domains, respectively, were included in the analysis, and each meaning unit was labeled and coded as problematic symptoms based on key attributes (physical, cognitive, social, and unspecified). Discrepancies in the mapping process were resolved by consensus between 2 senior investigators (CBF and ICH). PROMIS has applied rigorous standards to develop a comprehensive list of PRO items, therefore serving as a foundation for evaluating PRO contents [33-37]. This mapping process has been adopted in previous research to facilitate the abstraction and mapping of qualitative data [38-40]. In this study, the expert-labeled symptoms attributed to each meaning unit were deemed the gold standard for testing the validity of NLP/ML methods.

We evaluated the interrater reliability based on the raw concordance rate (defined as the percentage of coded meaning units that 2 coders provide concordant ratings), and Cohen κ statistic (defined as the number of concordant ratings to the number of discordant ratings while considering the agreement that is expected by chance). In our study, raw concordance rates were 88% for the pain interference domain and 86% for the fatigue domain. Cohen κ statistic was 0.6 for both domains, which is considered moderate or good reliability for coding qualitative PRO data [41].

NLP/ML Pipeline

Figure 1 outlines the pipeline of NLP/ML methods consisting of 2 key components: (1) extracting semantic features from the unstructured PROs and (2) using expert-labeled attributes of symptoms to validate NLP/ML–generated semantic features. We used the Word2vec [29] and BERT [30] methods to create multivariate semantic features (ie, word vectors) for each word from the meaning units. The BERT method embeds deep neural networks as a single step to perform abstraction and validation for the semantic features of symptom data simultaneously, whereas Word2vec/ML techniques involve 2 separate steps to achieve these tasks (Figure 1 and Multimedia Appendix 4).

Figure 1. The natural language processing and machine learning pipeline to analyze unstructured patient-reported outcomes data. BERT: bidirectional encoder representations from transformers; PROs: patient-reported outcomes; SVM: support vector machine; XGBoost: extreme gradient boosting.



BERT (Base, Uncased) for PRO Feature Extraction and Validation

The BERT (base, uncased; or the *BERT* hereafter), our primary interest in the NLP method, consists of the multilayer neural networks known as encoder transformers, and each generates context-dependent word features by weighting the features of each word with the other words in the meaning units [30,42]. We used 12 stacked layers of encoders to explore phrase-level, syntactic, semantic, and contextual information [42]. Specifically, we used the semantic features pretrained by articles published in BooksCorpus and Wikipedia to generate general word semantic meanings (pretrained model in Multimedia Appendix 5 [30,43]). The BERT model is augmented with a classification component, consisting of a feed-forward neural network and a softmax layer [44] to classify unstructured PROs (fine-tuning process in Multimedia Appendices 5 and 6). This augmented model was fine-tuned by the meaning units collected from interviews, which adapts the sentence contextual representation in encoders to the symptom-related contexts, and the parameters in the classification component were estimated simultaneously in one step.

Specifically, we used the pretrained model (BERT [base, uncased]) from the huggingface model repository, which was a pytorch implementation of the base BERT model [30]. The pretrained model is essentially based on the text passages included in BooksCorpus [43] and the English Wikipedia [30]. The weight parameters in the pretrained BERT model were further fine-tuned with the texts in the meaning units from our interview data when the BERT model was used for the downstream classification task of the meaning units through the BertForSequenceClassification object in the pytorch_transformers module. The use of BooksCorpus and Wikipedia is appropriate for our survivors of pediatric cancer as both contain comprehensive generic terms that capture the heterogeneous health status experienced by varying survivors of cancer, ranging from healthy (no late effects and no symptoms) to ill (severe late effects with severe symptoms).

Word2vec Method for PRO Feature Extraction and ML for Validation

We used Word2vec, our secondary interest in the NLP method, to extract semantic features based on the similarity of words in meaning units. Embedded with a one-level neural network model (Multimedia Appendix 7), Word2vec defines the semantic similarity across different words by using a specific word to search and connect other words nearby, given the hypothesis that a word's meaning is given by adjacent words [45,46]. We adopted the semantic features already pretrained by English articles from Wikipedia [47,48] to generate and fine-tune the semantic meanings of the meaning units through our data (Figure 1; Multimedia Appendices 6 and 7).

We used 2 ML methods, including the extreme gradient boosting (XGBoost) [25] and the SVM [24], to validate the semantic features derived from Word2vec in associations with the expert-labeled symptom attributes. ML modeling was used to account for high dimensional structures of semantic features created by Word2vec [29] (Multimedia Appendix 7). Specifically, XGBoost is a robust regression tree approach that includes multiple simple decision trees to iteratively refine the model performance by minimizing the difference between the expected and expert-labeled outcomes. In contrast, SVM is a classical ML algorithm that aims to find a decision boundary to separate the semantic features corresponding to the expert-labeled attributes by minimizing classification errors.

Alternative Methods for PRO Feature Extraction

In addition to the BERT, Word2vec/SVM, and Word2vec/XGBoost models, we conducted pilot analyses to evaluate 6 alternative NLP/ML models, including the TF-IDF/SVM, GloVe/SVM, and GloVe/XGBoost, as well as 3 extended BERT models (BioBERT, BlueBERT, and Clinical BERT). Briefly, the TF-IDF is an automatic text analysis that accounts for the number of times a word appears in a document and the number of documents that contain the word [27]. The GloVe method identifies the global word similarity over several meaning units (ie, our unit of analysis) or the entire interview [28]. The 3 alternative BERT models for pilot testing included the BioBERT (base, cased and trained on PubMed 1M) [49], BlueBERT (base, uncased and trained on PubMed) [50], and Clinical BERT (base, cased, initialized from BioBERT and trained on all MIMIC-III notes) [51].

As demonstrated in Multimedia Appendix 8, the areas under the precision-recall (PR) curves for the BERT model were significantly superior to the TF-IDF/SVM, GloVe/SVM, and GloVe/XGBoost (all attributes over 2 symptom domains) and were significantly superior to the BioBERT, BlueBERT, and Clinical BERT models (especially physical and cognitive attributes in the pain interference domain). In addition, the use of GloVe/SVM, Word2vec/SVM, and Word2vec/XGBoost methods resulted in statistically nonsignificant differences. Model performances based on other evaluation metrics were reported in Multimedia Appendices 9 and 10. As the main purpose of this study was to identify the NLP/ML model with optimal performance for symptom assessment, we focused on comparisons between the BERT model (as a theoretically optimal method) and the Word2vec model accompanied by SVM and XGBoost (as a suboptimal method).

Model Training and Evaluation

We used a 5-folder nested cross-validation approach (Multimedia Appendix 11) to address the issue of small sample size, including the components of partitioning the training, validation and test sets, determining the tuning parameters in ML methods, and generating validation results. Given the 4-attribute classification (physical, cognitive, social, and unclassified) on each meaning unit, we used a one-versus-rest binary classifier to classify one attribute (physical, cognitive, or social) versus the remaining attributes (the reference) for model training and evaluation [52].

We used standard metrics to test the validity of NLP/ML models, including precision (ie, positive predictive value), sensitivity (ie, recall), specificity, accuracy (summarizing true positive and true negative), F1 score (summarizing sensitivity and positive predictive values), areas under the receiver operating characteristic (ROC) curve, and areas under the PR curve. In the case of imbalanced data (ie, a limited number of meaning units labeled as attribute presence versus that of the reference), the PR curve is more suitable than the ROC curve as the former focuses on precision and sensitivity related to true positive cases [53]. On the basis of a recommendation [53], we determined the baseline threshold for each attribute of a symptom domain as the percentage of meaning units that were rated by 2 coders or content experts (ie, the gold standard for labeling true presence of attribute), which represents the precision of a random guess classifier.

Our NLP framework benefits from the transfer learning framework, which uses a huge amount of related data in the public domains to improve the ML application with regular sample sizes. Specifically, our Word2vec and BERT models or algorithms were pretrained by millions of health-related information in the public domains (eg, Wikipedia). Our meaning units were only used to fine-tune or improve the pretrained model and as predictive samples. Although our sample size was not large, it was sufficient to achieve robust validation and predictive performance. The codes used for BERT modeling are available on the GitHub website [54]; the fully deidentified unstructured PRO data used in this study can be shared for research purposes on user's request.

Results

Participant Characteristics

Table 1 reports the participant characteristics. The mean (SD) ages of survivors (N=52) and caregivers (N=35) at interviews were 13.8 (2.8) and 39.6 (7.0) years, respectively. Approximately 42% (22/52) of survivors were treated for noncentral nervous system solid tumors and 33% (17/52) for leukemia. For meaning units, 391 in the pain interference domain—of the 391 units, 255 (65.2%) were from survivors, and 136 (34.8%) were from caregivers—and 423 in the fatigue domain—of the 423 units, 275 (65%) were from survivors, and 148 (35%) were from caregivers— were labeled and analyzed accordingly (Multimedia Appendix 12).

Table 1. Characteristics of study participants (N=87).

Characteristics	Survivors (n=52)	Caregivers (n=35)
Age at evaluation (years), mean (SD)	13.8 (2.8)	39.6 (7.0)
Sex, n (%)		
Female	31 (61)	32 (91)
Male	20 (39)	3 (9)
Race or ethnicity, n (%)		
White, non-Hispanic	30 (59)	24 (69)
Black, non-Hispanic	14 (28)	10 (29)
Other	7 (14)	1 (3.0)
Cancer diagnosis, n (%)		
Non-CNS[a] solid tumor	22 (42)	N/A[b]
Leukemia	17 (33)	N/A
CNS malignancy	9 (17)	N/A
Lymphoma	4 (8.0)	N/A

[a]CNS: central nervous system.

[b]N/A: not applicable.

Sensitivity, Specificity, Precision, and Accuracy for Pain Interference

Table 2 reports the model performance for the pain interference domain based on survivor and caregiver data. For the sensitivity metric, compared with Word2vec/SVM and Word2vec/XGBoost, BERT generated higher values in identifying problems with 3 attributes (physical, cognitive, and social); however, the values were largely <0.6. In contrast, all 3 methods produced specificity of >0.9, and Word2vec/XGBoost produced higher values in identifying problems with 3 attributes compared with BERT and Word2vec/SVM. For F1-statistics, BERT yielded higher values for all 3 attributes compared with Word2vec/SVM and Word2vec/XGBoost. BERT yielded higher accuracy for all 3 attributes compared with Word2vec/SVM and Word2vec/XGBoost; the values were all >0.8, specifically 0.931 (95% CI 0.905-0.957), 0.916 (95% CI 0.887-0.941), and 0.870 (95% CI 0.836-0.903) for cognitive, social, and physical attributes, respectively.

Table 2. Performance of natural language processing/machine learning models for pain interference domain by 3 symptom attributes.

Attributes and models	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1 (95% CI)	AUROCC[a] (95% CI)	AUPRC[b] (95% CI)
Physical							
BERT[c]	0.692 (0.555-0.811)	0.507 (0.387-0.618)	0.950 (0.924-0.972)	0.870 (0.836-0.903)	0.585 (0.467-0.683)	0.875 (0.824-0.948)	0.677 (0.568-0.770)
Word2vec/SVM[d]	0.722 (0.562-0.867)	0.366 (0.262-0.479)	0.969 (0.948-0.987)	0.859 (0.824-0.893)	0.486 (0.362-0.594)	0.868 (0.826-0.922)	0.623 (0.5090.743)
Word2vec/XGBoost[e]	0.697 (0.528-0.857)	0.324 (0.221-0.435)	0.969 (0.949-0.987)	0.852 (0.813-0.887)	0.442 (0.318-0.551)	0.830 (0.769-0.888)	0.553 (0.437-0.659)
Cognitive							
BERT	0.800 (0.657-0.935)	0.583 (0.432-0.735)	0.980 (0.964-0.994)	0.931 (0.905-0.957)	0.675 (0.543-0.779)	0.923 (0.879-0.997)	0.818 (0.735-0.917)
Word2vec/SVM	0.760 (0.583-0.920)	0.396 (0.254-0.533)	0.983 (0.967-0.994)	0.910 (0.882-0.939)	0.521 (0.361-0.648)	0.900 (0.863-0.957)	0.609 (0.434-0.761)
Word2vec/XGBoost	0.769 (0.500-1.000)	0.208 (0.104-0.333)	0.991 (0.980-1.000)	0.895 (0.867-0.926)	0.328 (0.178-0.474)	0.828 (0.748-0.905)	0.474 (0.321-0.630)
Social							
BERT	0.636 (0.461-0.800)	0.500 (0.349-0.652)	0.966 (0.946-0.983)	0.916 (0.887-0.941)	0.560 (0.410-0.690)	0.857 (0.786-0.918)	0.566 (0.402-0.750)
Word2vec/SVM	0.286 (0-0.668)	0.048 (0-0.118)	0.986 (0.973-0.997)	0.885 (0.854-0.916)	0.082 (0.035-0.200)	0.804 (0.742-0.878)	0.309 (0.173-0.426)
Word2vec/XGBoost	0.556 (0.222-0.875)	0.119 (0.029-0.229)	0.989 (0.977-0.997)	0.895 (0.864-0.923)	0.196 (0.072-0.343)	0.786 (0.728-0.850)	0.304 (0.148-0.420)

[a]AUROCC: area under the receiver operating characteristic curve.

[b]AUPRC: area under precision-recall curve.

[c]BERT: bidirectional encoder representations from transformers.

[d]SVM: support vector machine.

[e]XGBoost: extreme gradient boosting.

Sensitivity, Specificity, Precision, and Accuracy for Fatigue

Table 3 reports the model performance for the fatigue domain based on the survivor and caregiver data. For sensitivity, the BERT method generated higher values in identifying problems with 3 attributes compared with Word2vec/SVM and Word2vec/XGBoost; however, the values were largely <0.5, except cognitive attributes (0.757). In contrast, all 3 methods produced specificity >0.9, and Word2vec/SVM produced higher values in identifying problems with 3 attributes compared with BERT and Word2vec/XGBoost. The BERT model yielded higher F1-statistics for all 3 individual attributes compared with Word2vec/SVM and Word2vec/XGBoost. In addition, the BERT model produced higher accuracy for all 3 attributes compared with Word2vec/SVM and Word2vec/XGBoost; the values were all >0.8, specifically 0.929 (95% CI 0.903-0.953), 0.917 (95% CI 0.891-0.943), and 0.832 (95% CI 0.794-0.867) for cognitive, social, and physical attributes, respectively.

Table 3. Performance of natural language processing/machine learning models for fatigue domain by 3 symptom attributes.

Attributes and models	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1 (95% CI)	AUROCC[a] (95% CI)	AUPRC[b] (95% CI)
Physical							
BERT[c]	0.593 (0.468-0.717)	0.427 (0.315-0.538)	0.929 (0.901-0.956)	0.832 (0.794-0.867)	0.496 (0.384-0.593)	0.775 (0.723-0.848)	0.537 (0.443-0.634)
Word2vec/SVM[d]	0.600 (0.286-0.900)	0.073 (0.026-0.136)	0.988 (0.974-0.997)	0.810 (0.770-0.848)	0.130 (0.048-0.227)	0.726 (0.670-0.780)	0.375 (0.224-0.474)
Word2vec/XGBoost[e]	0.595 (0.432-0.773)	0.268 (0.169-0.364)	0.956 (0.934-0.977)	0.822 (0.784-0.858)	0.370 (0.250-0.474)	0.726 (0.665-0.798)	0.461 (0.338-0.575)
Cognitive							
BERT	0.803 (0.696-0.895)	0.757 (0.652-0.854)	0.963 (0.941-0.981)	0.929 (0.903-0.953)	0.779 (0.697-0.855)	0.948 (0.922-0.979)	0.855 (0.791-0.930)
Word2vec/SVM	0.829 (0.690-0.946)	0.414 (0.292-0.535)	0.983 (0.968-0.994)	0.889 (0.861-0.917)	0.552 (0.418-0.657)	0.917 (0.886-0.951)	0.730 (0.632-0.855)
Word2vec/XGBoost	0.767 (0.625-0.884)	0.471 (0.359-0.586)	0.972 (0.953-0.988)	0.889 (0.858-0.917)	0.584 (0.468-0.684)	0.860 (0.817-0.924)	0.659 (0.550-0.782)
Social							
BERT	0.679 (0.500-0.848)	0.422 (0.289-0.568)	0.976 (0.960-0.990)	0.917 (0.891-0.943)	0.521 (0.379-0.658)	0.796 (0.704-0.912)	0.561 (0.434-0.741)
Word2vec/SVM	0.778 (0.429-1.000)	0.156 (0.057-0.267)	0.995 (0.987-1.000)	0.905 (0.877-0.929)	0.259 (0.102-0.406)	0.817 (0.756-0.881)	0.393 (0.203-0.534)
Word2vec/XGBoost	0.571 (0.286-0.833)	0.178 (0.068-0.300)	0.984 (0.971-0.995)	0.898 (0.868-0.924)	0.271 (0.118-0.415)	0.780 (0.706-0.850)	0.330 (0.154-0.436)

[a]AUROCC: area under the receiver operating characteristic curve.

[b]AUPRC: area under precision-recall curve.

[c]BERT: bidirectional encoder representations from transformers.

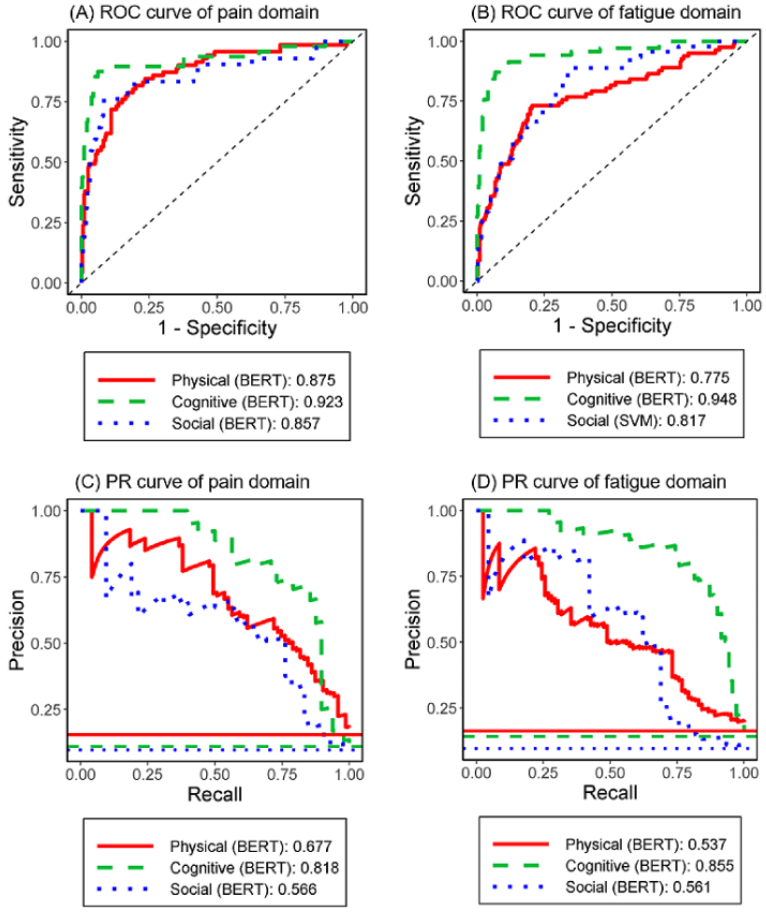
[d]SVM: support vector machine.

[e]XGBoost: extreme gradient boosting.

Area Under the ROC Curves for Pain Interference and Fatigue

Figure 2 (upper) displays the specific NLP/ML method that had the highest area under the ROC curves for each attribute (detailed results in Tables 2 and 3). The diagonal line represents the random guess (ie, reference). For the pain interference domain (left panel), the BERT model was superior to the Word2vec/SVM and Word2vec/XGBoost models, and the areas under the ROC curve were 0.923 (95% CI 0.879-0.997) for cognitive, 0.875 (95% CI 0.824-0.948) for physical attributes, and 0.857 (95% CI 0.786-0.918) for social attributes. For the fatigue domain (right panel), the BERT model was superior to the Word2vec/SVM and Word2vec/XGBoost models, and areas under the ROC curve were (0.948, 95% CI 0.922-0.979) for cognitive and 0.775 (95% CI 0.723-0.848) for physical attributes. The values of BERT were significantly higher in identifying problems with cognitive attributes in both pain interference and fatigue domains compared with Word2vec/XGBoost (P<.05; Multimedia Appendix 13).

Figure 2. Area under the receiver operating characteristic curves and precision-recall curves for the best models of pain interference domain (left column) and fatigue domain (right column) by 3 symptom attributes. BERT: bidirectional encoder representations from transformers; PR: precision recall; ROC: receiver operating characteristic; SVM: support vector machine.



Area Under the PR Curves for Pain Interference and Fatigue

Figure 2 (lower) displays the specific NLP/ML method that had the highest area under the PR curves for each attribute (see detailed results in Tables 2 and 3). The horizontal line at the bottom represents a random guess (ie, reference). For the pain interference domain (left panel), the BERT model was superior to the Word2vec/SVM and Word2vec/XGBoost models, and the areas under the PR curve were 0.818 (95% CI 0.735-0.917) for cognitive, 0.677 (95% CI 0.568-0.770) for physical attributes, and 0.566 (95% CI 0.402-0.750) for social attributes. For the fatigue domain (right panel), the BERT models were superior to the Word2vec/SVM and Word2vec/XGBoost models, and areas under the PR curve were 0.855 (95% CI 0.791-0.930) for cognitive, 0.561 (95% CI 0.434-0.741) for social attributes, and 0.537 (95% CI 0.443-0.634) for physical attributes. In addition, the values of BERT were significantly higher in identifying problems with cognitive and social attributes in both pain interference and fatigue domains compared with both Word2vec/SVM and Word2vec/XGBoost (P<.05; Multimedia Appendices 13 and 14).

Discussion

Principal Findings

Very limited studies have demonstrated the feasibility of applying NLP/ML methods to extract semantic features from unstructured PROs. This study applied different NLP/ML models to analyze PRO assessment in pediatric cancer survivorship, with a special focus on young survivors of pediatric cancer aged <18 years as a vulnerable population, and used rigorous methods to validate the performance of NLP/ML models. The results suggest that the BERT method outperformed the Word2vec/ML methods across different validation metrics in both the physical interference and fatigue symptom domains. Specifically, the BERT method yielded higher accuracy (>0.8), larger area under the ROC curve (>0.8, except for the social attribute in fatigue domain), and a larger area under the PR curve in identifying problems with all 3 attributes over 2 symptom domains compared with the Word2vec/SVM and Word2vec/XGBoost methods. The models with higher accuracy were characterized by high specificity (>0.9) but low sensitivity (<0.5) for all 3 attributes and 2 symptom domains.

The findings of high specificity and low sensitivity suggest that our NLP/ML algorithms can be used to identify problematic symptoms (ie, diagnostic confirmation) rather than for symptom screening. However, if the default threshold (ie, 0.5) for ROC

curves was changed to a lower value that mimics the proportion of meaning units labeled as the presence of the problematic attribute, both specificity and sensitivity will reach the level of 0.7-0.8. How to use NLP/ML techniques to convert unstructured PROs into semantic features and transform the data into meaningful diagnostic information for clinical decision-making is an emerging topic [20,55]. It is important to extend our NLP/ML pipeline to assess other aspects of symptom problems (eg, severity and interference) for cancer populations and in a longitudinal context, which is valuable for detecting changes in symptom patterns and identifying early signs of adverse events [22,56,57].

Comparisons of Model Performance

In both symptom domains, the performance of NLP/ML techniques (accuracy, F1 value, and areas under ROC and PR curves) in identifying problems with cognitive attributes was superior to physical and social attributes. Interestingly, model validity based on data collected from survivors and caregivers was slightly better than that of survivors alone (Multimedia Appendices 15 and 16). This finding is in part because of the inclusion of complementary information from survivors and caregivers and the increase in sample size.

The superior performance of NLP/ML techniques suggests the usefulness of interview-based methods for collecting unstructured PRO data to complement the survey-based methods that contain a prespecified fixed content of PROs in follow-up care among survivors of cancer. Using our validated NLP/ML algorithms to automatically abstract and label the semantic features of unstructured PROs derived from interviews represents an efficient strategy for collecting PRO data from busy clinics. Our NLP/ML approach can be extended to analyze other forms of unstructured PROs (eg, documented patient-clinician conversations and medical notes in EMRs) when data are available. Other novel technologies (eg, audio-recorded PROs) also deserve investigation in analyzing unstructured PROs. Multimodal sentiment analysis [58], which investigates affective states by extracting textual and audio features, can be combined with the semantic features from NLP to obtain a comprehensive understanding of survivors' PROs. The successful application of NLP/ML for PRO assessment ideally requires the implementation of integrated platforms that interconnect the EHR-based medical note systems, NLP/ML analytics, and supportive tools for result display, clinical interpretation, and treatment recommendation [20,59-61]. The integrated platforms will facilitate clinicians in clinical decision-making for caring for survivors of cancer whose complex late medical effects can be predicted by the deterioration of symptoms and clinical parameters.

The superior performance of BERT to the Word2vec/ML method is because of the flexible design of BERT that accounts for contextual information of PROs. Basically, BERT includes multilayer deep neural networks (illustrated in self-attention layers of the fine-tuning process; Multimedia Appendix 5) to enable flexible feature extraction at different levels, such as syntactic, semantic, and contextual information. In comparison, Word2vec includes a one-level shallow neural network with limited flexibility. Uniquely, the semantic features derived by BERT capture different meanings of the same word in different contexts, whereas Word2vec generates static semantic features for each word that does not vary in different contexts.

Different NLP Methods for Analyzing Unstructured PRO Data

The clinical application of NLP/ML in PRO research is still in its infancy. This study used the BERT model pretrained by Wikipedia and BooksCorpus to generate general semantic features as a starting point. The use of BooksCorpus and Wikipedia is appropriate for survivors of pediatric cancer, resulting in satisfactory model performance. This is because BooksCorpus and Wikipedia contain comprehensive generic terms that capture the heterogeneous health conditions experienced by various populations, including survivors of cancer, ranging from healthy (no late effects and no symptoms) to ill (severe late effects with severe symptoms). Alternatively, BERT models can be pretrained using larger free text data to generate comprehensive features of PROs. Similar methods may include SciBERT [62], trained by texts in Semantic Scholar; BioBERT [49], trained by texts in PubMed; and Clinical BERT [51], trained by clinical notes in MIMIC-III [63]. In addition, the health knowledge graph [64] can be used to integrate different concepts from various data elements in multiomics frameworks (including unstructured PROs in medical notes, structured PROs from patient survey, imaging, genetics, and treatment profiles), and analyze complex relationships among these data to improve evaluations of survivorship outcomes through a multitask learning framework [65].

Limitations

This study contains several limitations. First, our samples were limited to survivors of pediatric cancer who were treated at a single institution. However, our samples represent diverse diagnoses, ages, races and ethnicities, and families residing in counties with poverty levels similar to the national average [15]. Second, we only analyzed pain interference and fatigue domains and restricted them to 3 key attributes of symptoms. Future studies are encouraged to apply our NLP/ML pipeline to analyze other PRO domains and include more comprehensive attribute classifications. Third, our data were collected cross-sectionally, which merely provides a snapshot of PROs. Future studies are needed to test the validity of abstracting longitudinal unstructured PROs to identify time-dependent patterns. In summary, we demonstrated a robust validity of NLP/ML algorithms in abstracting and analyzing unstructured PROs collected from interviews with childhood survivors of cancer and caregivers. These promising results suggest the utility of NLP/ML methods in future works for monitoring survivors' PROs and the opportunity of extending our methods to other PRO domains and data collection systems (eg, audio-recorded or medical notes) under a unified platform that integrates EHR-based data collection systems, NLP/ML analytics, and supportive tools for interpretation of results and treatment recommendations. Integration of NLP/ML-based PRO assessment to complement other clinical data will facilitate the improvement of follow-up care for survivors of cancer.

Acknowledgments

The authors thank Rachel M Keesey and Ruth J Eliason for conducting in-depth interviews with study participants, and Jennifer L Clegg and Conor M Jones for labeling the attributes of symptoms based on the interview data.

Authors' Contributions

ZL, JAS, and ICH contributed to the concept and design; CBF provided administrative support; MMH and LLR contributed to the provision of study materials JNB and ICH contributed to collection and assembly of data; ZL, JAS, JXW, and ICH contributed to data analysis and interpretation; ZL, JAS, and ICH contributed to manuscript writing; and all authors contributed to editing and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guides for pain interference domain (cancer survivor).
[DOCX File , 26 KB - jmir_v23i11e26777_app1.docx]

Multimedia Appendix 2
Interview guides for fatigue domain (cancer survivor).
[DOCX File , 26 KB - jmir_v23i11e26777_app2.docx]

Multimedia Appendix 3
Examples of meaning units derived from study participants and corresponding attributes.
[DOCX File , 24 KB - jmir_v23i11e26777_app3.docx]

Multimedia Appendix 4
The natural language processing/machine learning pipeline to analyze unstructured patient-reported outcomes data.
[DOCX File , 88 KB - jmir_v23i11e26777_app4.docx]

Multimedia Appendix 5
Concept of bidirectional encoder representations from transformers [base, uncased] techniques.
[DOCX File , 489 KB - jmir_v23i11e26777_app5.docx]

Multimedia Appendix 6
Tools or packages and fine-tuned hyper-parameters to analyze natural language processing/machine learning models.
[DOCX File , 24 KB - jmir_v23i11e26777_app6.docx]

Multimedia Appendix 7
Concept of Word2Vec techniques.
[DOCX File , 426 KB - jmir_v23i11e26777_app7.docx]

Multimedia Appendix 8
The changes of the area under the precision-recall curve among different natural language processing/machine learning models.
[DOCX File , 26 KB - jmir_v23i11e26777_app8.docx]

Multimedia Appendix 9
Performance of natural language processing/machine learning models for pain interference domain by three symptom attributes (cancer survivors and caregivers).
[DOCX File , 30 KB - jmir_v23i11e26777_app9.docx]

Multimedia Appendix 10
Performance of natural language processing/machine learning models for fatigue domain by three symptom attributes (cancer survivors and caregivers).

[DOCX File , 30 KB - jmir_v23i11e26777_app10.docx]

Multimedia Appendix 11
Five-fold cross-validation methods.
[DOCX File , 767 KB - jmir_v23i11e26777_app11.docx]

Multimedia Appendix 12
Frequency of attributes in pain interference and fatigue domains labeled by content experts.
[DOCX File , 24 KB - jmir_v23i11e26777_app12.docx]

Multimedia Appendix 13
The changes in the area under the receiver operating characteristic curve and precision-recall curve among different natural language processing/machine learning models (survivors and caregivers).
[DOCX File , 26 KB - jmir_v23i11e26777_app13.docx]

Multimedia Appendix 14
Precision-recall curves for pain interference and fatigue domains by three symptom attributes (survivors and caregivers).
[DOCX File , 153 KB - jmir_v23i11e26777_app14.docx]

Multimedia Appendix 15
Performance of natural language processing/machine learning models for pain interference domain by three symptom attributes (survivors only).
[DOCX File , 26 KB - jmir_v23i11e26777_app15.docx]

Multimedia Appendix 16
Performance of natural language processing/machine learning models for fatigue domain by three symptom attributes (survivors only).
[DOCX File , 26 KB - jmir_v23i11e26777_app16.docx]

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics. CA Cancer J Clin 2019 Sep;69(5):363-385 [FREE Full text] [doi: 10.3322/caac.21565] [Medline: 31184787]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019 Jan;69(1):7-34 [FREE Full text] [doi: 10.3322/caac.21551] [Medline: 30620402]
3. Phillips SM, Padgett LS, Leisenring WM, Stratton KK, Bishop K, Krull KR, et al. Survivors of childhood cancer in the United States: Prevalence and burden of morbidity. Cancer Epidemiol Biomarkers Prev 2015 Apr;24(4):653-663 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-1418] [Medline: 25834148]
4. Oeffinger KC, Mertens AC, Sklar CA, Kawashima T, Hudson MM, Meadows AT, Childhood Cancer Survivor Study. Chronic health conditions in adult survivors of childhood cancer. N Engl J Med 2006 Oct 12;355(15):1572-1582. [doi: 10.1056/NEJMsa060185] [Medline: 17035650]
5. Hudson MM, Ness KK, Gurney JG, Mulrooney DA, Chemaitilly W, Krull KR, et al. Clinical ascertainment of health outcomes among adults treated for childhood cancer. JAMA 2013 Jun 12;309(22):2371-2381 [FREE Full text] [doi: 10.1001/jama.2013.6296] [Medline: 23757085]
6. Armstrong GT, Chen Y, Yasui Y, Leisenring W, Gibson TM, Mertens AC, et al. Reduction in late mortality among 5-year survivors of childhood cancer. N Engl J Med 2016 Mar 03;374(9):833-842 [FREE Full text] [doi: 10.1056/NEJMoa1510795] [Medline: 26761625]
7. Huang I, Brinkman TM, Kenzik K, Gurney JG, Ness KK, Lanctot J, et al. Association between the prevalence of symptoms and health-related quality of life in adult survivors of childhood cancer: a report from the st jude lifetime cohort study. J Clin Oncol 2013 Nov 20;31(33):4242-4251 [FREE Full text] [doi: 10.1200/JCO.2012.47.8867] [Medline: 24127449]
8. Zeltzer LK, Recklitis C, Buchbinder D, Zebrack B, Casillas J, Tsao JC, et al. Psychological status in childhood cancer survivors: a report from the childhood cancer survivor study. J Clin Oncol 2009 May 10;27(14):2396-2404 [FREE Full text] [doi: 10.1200/JCO.2008.21.1433] [Medline: 19255309]
9. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: Life-long risks and responsibilities. Nat Rev Cancer 2014 Jan;14(1):61-70 [FREE Full text] [doi: 10.1038/nrc3634] [Medline: 24304873]
10. Spathis A, Booth S, Grove S, Hatcher H, Kuhn I, Barclay S. Teenage and young adult cancer-related fatigue is prevalent, distressing, and neglected: It is time to intervene. A systematic literature review and narrative synthesis. J Adolesc Young Adult Oncol 2015 Mar;4(1):3-17 [FREE Full text] [doi: 10.1089/jayao.2014.0023] [Medline: 25852970]

11. Rosenberg AR, Orellana L, Ullrich C, Kang T, Geyer JR, Feudtner C, et al. Quality of life in children with advanced cancer: a report from the PediQUEST study. J Pain Symptom Manage 2016 Aug;52(2):243-253 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.04.002] [Medline: 27220948]
12. Wolfe J, Orellana L, Ullrich C, Cook EF, Kang TI, Rosenberg A, et al. Symptoms and distress in children with advanced cancer: Prospective patient-reported outcomes from the PediQUEST study. J Clin Oncol 2015 Jun 10;33(17):1928-1935 [FREE Full text] [doi: 10.1200/JCO.2014.59.1222] [Medline: 25918277]
13. van Deuren S, Boonstra A, van Dulmen-den Broeder E, Blijlevens N, Knoop H, Loonen J. Severe fatigue after treatment for childhood cancer. Cochrane Database Syst Rev 2020 Mar 03;3(3):CD012681 [FREE Full text] [doi: 10.1002/14651858.CD012681.pub2] [Medline: 32124971]
14. Alberts NM, Gagnon MM, Stinson JN. Chronic pain in survivors of childhood cancer: a developmental model of pain across the cancer trajectory. Pain 2018 Oct;159(10):1916-1927. [doi: 10.1097/j.pain.0000000000001261] [Medline: 29708940]
15. Pierzynski JA, Clegg JL, Sim J, Forrest CB, Robison LL, Hudson MM, et al. Patient-reported outcomes in paediatric cancer survivorship: a qualitative study to elicit the content from cancer survivors and caregivers. BMJ Open 2020 May 17;10(5):e032414 [FREE Full text] [doi: 10.1136/bmjopen-2019-032414] [Medline: 32423926]
16. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. CA Cancer J Clin 2007;57(5):278-300 [FREE Full text] [doi: 10.3322/CA.57.5.278] [Medline: 17855485]
17. Warsame R, D'Souza A. Patient reported outcomes have arrived: a practical overview for clinicians in using patient reported outcomes in oncology. Mayo Clin Proc 2019 Nov;94(11):2291-2301 [FREE Full text] [doi: 10.1016/j.mayocp.2019.04.005] [Medline: 31563425]
18. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. Yearb Med Inform 2017 Aug;26(1):214-227 [FREE Full text] [doi: 10.15265/IY-2017-029] [Medline: 29063568]
19. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
20. Forsyth AW, Barzilay R, Hughes KS, Lui D, Lorenz KA, Enzinger A, et al. Machine learning methods to extract documentation of breast cancer symptoms from electronic health records. J Pain Symptom Manage 2018 Jun;55(6):1492-1499 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.02.016] [Medline: 29496537]
21. Hyun S, Johnson SB, Bakken S. Exploring the ability of natural language processing to extract data from nursing narratives. Comput Inform Nurs 2009;27(4):215-223 [FREE Full text] [doi: 10.1097/NCN.0b013e3181a91b58] [Medline: 19574746]
22. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc 2019 Apr 01;26(4):364-379 [FREE Full text] [doi: 10.1093/jamia/ocy173] [Medline: 30726935]
23. Chandran D, Robbins DA, Chang C, Shetty H, Sanyal J, Downs J, et al. Use of natural language processing to identify obsessive compulsive symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. Sci Rep 2019 Oct 02;9(1):14146 [FREE Full text] [doi: 10.1038/s41598-019-49165-2] [Medline: 31578348]
24. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, England, UK: Cambridge University Press; 2000.
25. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: Association for Computing Machinery; 2016 Aug Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2016; San Francisco, California, USA p. 785-794. [doi: 10.1145/2939672.2939785]
26. Wong J, Horwitz MM, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. Curr Epidemiol Rep 2018 Dec;5(4):331-342 [FREE Full text] [doi: 10.1007/s40471-018-0165-9] [Medline: 30555773]
27. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag 1988;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]
28. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).: Association for Computational Linguistics; 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar p. A. [doi: 10.3115/v1/d14-1162]
29. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems. 2013 Presented at: 26th International Conference on Neural Information Processing Systems; December 5 - 10, 2013; Lake Tahoe Nevada p. 3111-3119 URL: https://dl.acm.org/doi/10.5555/2999792.2999959
30. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2018; New Orleans, Louisiana p. 4171-4186. [doi: 10.18653/v1/n18-2]

31. Forrest CB, Forrest KD, Clegg JL, de la Motte A, Amaral S, Grossman AB, et al. Establishing the content validity of PROMIS pediatric pain interference, fatigue, sleep disturbance, and sleep-related impairment measures in children with chronic kidney disease and crohn's disease. J Patient Rep Outcomes 2020 Feb 12;4(1):11 [FREE Full text] [doi: 10.1186/s41687-020-0178-2] [Medline: 32052205]
32. PROMIS measures for pediatric self-report (ages 8-17) and parent proxy report (ages 5-17). Health Measures. URL: https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-pediatric-measures [accessed 2021-09-09]
33. Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai J, et al. PROMIS pediatric pain interference scale: An item response theory analysis of the pediatric pain item bank. J Pain 2010 Nov;11(11):1109-1119 [FREE Full text] [doi: 10.1016/j.jpain.2010.02.005] [Medline: 20627819]
34. Lai J, Stucky BD, Thissen D, Varni JW, DeWitt EM, Irwin DE, et al. Development and psychometric properties of the PROMIS(®) pediatric fatigue item banks. Qual Life Res 2013 Nov;22(9):2417-2427 [FREE Full text] [doi: 10.1007/s11136-013-0357-1] [Medline: 23378106]
35. Bevans KB, Gardner W, Pajer KA, Becker B, Carle A, Tucker CA, et al. Psychometric evaluation of the PROMIS® pediatric psychological and physical stress experiences measures. J Pediatr Psychol 2018 Jul 01;43(6):678-692 [FREE Full text] [doi: 10.1093/jpepsy/jsy010] [Medline: 29490050]
36. Lai J, Nowinski CJ, Zelko F, Wortman K, Burns J, Nordli DR, et al. Validation of the Neuro-QoL measurement system in children with epilepsy. Epilepsy Behav 2015 May;46:209-214 [FREE Full text] [doi: 10.1016/j.yebeh.2015.02.038] [Medline: 25862469]
37. Ravens-Sieberer U, Devine J, Bevans K, Riley AW, Moon J, Salsman JM, et al. Subjective well-being measures for children were developed within the PROMIS project: Presentation of first results. J Clin Epidemiol 2014 Feb;67(2):207-218 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.018] [Medline: 24295987]
38. Ibragimova N, Pless M, Adolfsson M, Granlund M, Björck-Åkesson E. Using content analysis to link texts on assessment and intervention to the international classification of functioning, disability and health - version for children and youth (ICF-CY). J Rehabil Med 2011 Jul;43(8):728-733 [FREE Full text] [doi: 10.2340/16501977-0831] [Medline: 21732007]
39. Hwang A, Yen C, Liou T, Bedell G, Granlund M, Teng S, et al. Development and validation of the ICF-CY-based functioning scale of the disability evaluation system--child version in Taiwan. J Formos Med Assoc 2015 Dec;114(12):1170-1180 [FREE Full text] [doi: 10.1016/j.jfma.2015.11.002] [Medline: 26705138]
40. Güeita-Rodríguez J, Florencio LL, Arias-Buría JL, Lambeck J, Fernández-de-Las-Peñas C, Palacios-Ceña D. Content comparison of aquatic therapy outcome measures for children with neuromuscular and neurodevelopmental disorders using the international classification of functioning, disability, and health. Int J Environ Res Public Health 2019 Nov 02;16(21):4263 [FREE Full text] [doi: 10.3390/ijerph16214263] [Medline: 31684043]
41. Hahn EA, Cella D, Chassany O, Fairclough DL, Wong GY, Hays RD, Clinical Significance Consensus Meeting Group. Precision of health-related quality-of-life data compared with other clinical measures. Mayo Clin Proc 2007 Oct;82(10):1244-1254. [doi: 10.4065/82.10.1244] [Medline: 17908530]
42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. Advances in Neural Information Processing Systems. 2017. URL: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf [accessed 2021-09-09]
43. Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015 Dec Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); Dec. 7-13, 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.11]
44. Goodfellow I, Yoshua B, Courville A. Deep Learning. Boston, Massachusetts, USA: MIT Press; 2016.
45. Harris ZS. Distributional Structure. Word 2015 Dec 04;10(2-3):146-162. [doi: 10.1080/00437956.1954.11659520]
46. Sadeghi Z, McClelland JL, Hoffman P. You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes. Neuropsychologia 2015 Sep;76:52-61 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2014.08.031] [Medline: 25196838]
47. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Trans Assoc Comput Linguist 2017 Dec;5:135-146. [doi: 10.1162/tacl_a_00051]
48. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T. Learning word vectors for 157 languages. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).: European Language Resources Association (ELRA); 2018 Presented at: Eleventh International Conference on Language Resources and Evaluation (LREC); May 2018; Miyazaki, Japan URL: https://aclanthology.org/L18-1550
49. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240. [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
50. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. In: Proceedings of the 18th BioNLP Workshop and Shared Task.: Association for Computational

Linguistics; 2019 Presented at: 18th BioNLP Workshop and Shared Task; August, 2019; Florence, Italy. [doi: 10.18653/v1/w19-5006]
51. Alsentzer E, Murphy J, Boag W, Weng W, Jindi D, Naumann T. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop.: Association for Computational Linguistics; 2019 Presented at: 2nd Clinical Natural Language Processing Workshop; June 2019; Minneapolis, Minnesota, USA. [doi: 10.18653/v1/w19-1909]
52. Bishop CM. Pattern Recognition and Machine Learning. New York, USA: Springer; 2006.
53. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015 Mar 04;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
54. Zhaohualu - nlp4pro1. GitHub. URL: https://github.com/zhaohualu/nlp4pro1 [accessed 2021-09-28]
55. Yim W, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. JAMA Oncol 2016 Jun 01;2(6):797-804. [doi: 10.1001/jamaoncol.2016.0213] [Medline: 27124593]
56. Heintzelman NH, Taylor RJ, Simonsen L, Lustig R, Anderko D, Haythornthwaite JA, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. J Am Med Inform Assoc 2013;20(5):898-905 [FREE Full text] [doi: 10.1136/amiajnl-2012-001076] [Medline: 23144336]
57. Jensen K, Soguero-Ruiz C, Oyvind MK, Lindsetmo R, Kouskoumvekaki I, Girolami M, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017 Apr 07;7:46226 [FREE Full text] [doi: 10.1038/srep46226] [Medline: 28387314]
58. Soleymani M, Garcia D, Jou B, Schuller B, Chang S, Pantic M. A survey of multimodal sentiment analysis. Image Vis Comput 2017 Sep;65:3-14. [doi: 10.1016/j.imavis.2017.08.003]
59. Wi C, Sohn S, Rolfes MC, Seabright A, Ryu E, Voge G, et al. Application of a natural language processing algorithm to asthma ascertainment: an automated chart review. Am J Respir Crit Care Med 2017 Aug 15;196(4):430-437 [FREE Full text] [doi: 10.1164/rccm.201610-2006OC] [Medline: 28375665]
60. Hong N, Wen A, Shen F, Sohn S, Wang C, Liu H, et al. Developing a scalable FHIR-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. JAMIA Open 2019 Oct 18;2(4):570-579 [FREE Full text] [doi: 10.1093/jamiaopen/ooz056] [Medline: 32025655]
61. Bozkurt S, Paul R, Coquet J, Sun R, Banerjee I, Brooks JD, et al. Phenotyping severity of patient-centered outcomes using clinical notes: A prostate cancer use case. Learn Health Syst 2020 Jul 17;4(4):e10237 [FREE Full text] [doi: 10.1002/lrh2.10237] [Medline: 33083539]
62. Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).: Association for Computational Linguistics; 2019 Presented at: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-1371]
63. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 04;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
64. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Sci Rep 2017 Jul 20;7(1):5994 [FREE Full text] [doi: 10.1038/s41598-017-05778-z] [Medline: 28729710]
65. Zhang Y, Yang Q. An overview of multi-task learning. Natl Sci Rev 2017 Sep 01;5(1):30-43. [doi: 10.1093/nsr/nwx105]

Abbreviations

BERT: bidirectional encoder representations from transformers
EMR: electronic medical record
ML: machine learning
NLP: natural language processing
PR: precision-recall
PRO: patient-reported outcome
PROMIS: Patient-Reported Outcomes Measurement Information System
ROC: receiver operating characteristic
SVM: support vector machine
TF-IDF: term frequency–inverse document frequency
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 25.12.20; peer-reviewed by K Reuter, T Basu; comments to author 03.02.21; revised version received 20.03.21; accepted 12.08.21; published 03.11.21.

Please cite as:
Lu Z, Sim JA, Wang JX, Forrest CB, Krull KR, Srivastava D, Hudson MM, Robison LL, Baker JN, Huang IC
Natural Language Processing and Machine Learning Methods to Characterize Unstructured Patient-Reported Outcomes: Validation Study
J Med Internet Res 2021;23(11):e26777
URL: https://www.jmir.org/2021/11/e26777
doi:10.2196/26777
PMID:34730546

Review

Use of Patient-Reported Outcome Measures and Patient-Reported Experience Measures Within Evaluation Studies of Telemedicine Applications: Systematic Review

Andreas Knapp[1], MSc; Lorenz Harst[1], MA; Stefan Hager[2], MPH; Jochen Schmitt[1], MPH, Prof Dr; Madlen Scheibe[1], Dipl Soc, Dr rer medic

[1]Center for Evidence-Based Healthcare, University Hospital Carl Gustav Carus, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

[2]Comprehensive Pain Center, University Hospital Carl Gustav Carus Dresden, Dresden, Germany

Corresponding Author:
Andreas Knapp, MSc
Center for Evidence-Based Healthcare
University Hospital Carl Gustav Carus, Carl Gustav Carus Faculty of Medicine
Technische Universität Dresden
Fetscherstrasse 74
Dresden, 01307
Germany
Phone: 49 3514585665
Email: andreas.knapp@uniklinikum-dresden.de

Abstract

Background: With the rise of digital health technologies and telemedicine, the need for evidence-based evaluation is growing. Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) are recommended as an essential part of the evaluation of telemedicine. For the first time, a systematic review has been conducted to investigate the use of PROMs and PREMs in the evaluation studies of telemedicine covering all application types and medical purposes.

Objective: This study investigates the following research questions: in which scenarios are PROMs and PREMs collected for evaluation purposes, which PROM and PREM outcome domains have been covered and how often, which outcome measurement instruments have been used and how often, does the selection and quantity of PROMs and PREMs differ between study types and application types, and has the use of PROMs and PREMs changed over time.

Methods: We conducted a systematic literature search of the MEDLINE and Embase databases and included studies published from inception until April 2, 2020. We included studies evaluating telemedicine with patients as the main users; these studies reported PROMs and PREMs within randomized controlled trials, controlled trials, noncontrolled trials, and feasibility trials in English and German.

Results: Of the identified 2671 studies, 303 (11.34%) were included; of the 303 studies, 67 (22.1%) were feasibility studies, 70 (23.1%) were noncontrolled trials, 20 (6.6%) were controlled trials, and 146 (48.2%) were randomized controlled trials. Health-related quality of life (n=310; mean 1.02, SD 1.05), emotional function (n=244; mean 0.81, SD 1.18), and adherence (n=103; mean 0.34, SD 0.53) were the most frequently assessed outcome domains. Self-developed PROMs were used in 21.4% (65/303) of the studies, and self-developed PREMs were used in 22.3% (68/303). PROMs (n=884) were assessed more frequently than PREMs (n=234). As the evidence level of the studies increased, the number of PROMs also increased (τ=−0.45), and the number of PREMs decreased (τ=0.35). Since 2000, not only has the number of studies using PROMs and PREMs increased, but the level of evidence and the number of outcome measurement instruments used have also increased, with the number of PREMs permanently remaining at a lower level.

Conclusions: There have been increasingly more studies, particularly high-evidence studies, which use PROMs and PREMs to evaluate telemedicine. PROMs have been used more frequently than PREMs. With the increasing maturity stage of telemedicine applications and higher evidence level, the use of PROMs increased in line with the recommendations of evaluation guidelines. Health-related quality of life and emotional function were measured in almost all the studies. Simultaneously, health literacy as a precondition for using the application adequately, alongside proper training and guidance, has rarely been reported. Further efforts should be pursued to standardize PROM and PREM collection in evaluation studies of telemedicine.

(J Med Internet Res 2021;23(11):e30042) doi:10.2196/30042

KEYWORDS

telemedicine; telehealth; evaluation; outcome; patient-reported outcome measures; patient-reported outcome; patient-reported experience measures; patient-reported experience; measurement instrument; questionnaire

Introduction

Background

With the rise of digital health technologies and telemedicine services, the need for evidence-based evaluation is growing [1]. Over the past years, several evaluation guidelines that address study types, outcomes, and patient perspectives, among other requirements have been published [2-7]. The two best-known and most commonly used evaluation guidelines are the *Model for Assessment of Telemedicine* (MAST) *applications* [2] and the *evidence standards framework for digital health technologies* of the English National Institute for Health and Care Excellence (NICE framework) [3]. They have been used in several evaluation studies over the years [1,8-10].

Focusing on outcomes, MAST provides the following elements as part of a multidisciplinary evaluation of telemedicine applications: clinical effectiveness, patient perspective, safety, economic aspects, organizational aspects, and sociocultural, ethical, and legal aspects [2]. The patient's perspective is evaluated by patient-reported outcome measures (PROMs), such as health-related quality of life (HRQoL) or behavioral outcomes, the latter being relevant when focusing on the domain of clinical effectiveness. In addition, patient-reported experience measures (PREMs) should be a part of the evaluation to assess satisfaction and acceptance, understanding of information, confidence in the treatment, ability to use the application, and empowerment [2]. The NICE framework provides minimum evidence standards and best practice standards for the evaluation of digital health technologies according to the degree of the treatment. Among them are, for example, the demonstration of effectiveness, use of behavior change techniques, and economic aspects. It also recommends the assessment of patient-centered outcomes in complex digital health technologies and specifically states that many of these outcomes should be measured using PROMs [3]. This demonstrates the importance of PROM and PREM in the context of evaluation studies of telemedicine applications.

The US Food and Drug Administration refers to PROMs as "any reports coming directly from patients about how they function or feel in relation to a health condition and its therapy, without interpretation of the patient's responses by a clinician, or anyone else" [11]. These reports are ideally collected using validated outcome measurement instruments (OMIs), which are regarded as cost-effective, efficient, and scalable, especially in the early stages of development of an innovative intervention [1]. In addition, PROMs are classified according to generic, disease-specific, and target group–specific OMIs [12].

OMIs that quantify the experience, satisfaction, acceptance, or quality of care from the patients' perspective are called PREMs. The goal of PREMs is to measure and report whether the provided care meets the expectations of the patients. Thus, PREMs are an indicator of patient centeredness and service quality in health care [13].

In the past, PROMs and PREMs have been used to evaluate the effectiveness and quality of care achieved when implementing telemedicine applications. Reviews of evaluation studies regarding telemedicine applications showed that single outcome domains such as HRQoL and psychological outcomes were used for specific use cases, such as inflammatory bowel disease management [14], adherence, self-efficacy, and self-management for medication management [15]. PREMs were used, for example, to measure satisfaction with knee pain management [16].

In summary, PROMs and PREMs have been recommended and already used for the evaluation of telemedicine applications. However, to the best of our knowledge, no systematic review exists to date that investigates the characteristics of the use of PROMs and PREMs in evaluation studies of telemedicine applications irrespective of application type and medical purpose.

It is still not known which and how often outcome domains and OMIs have been used in evaluation studies and whether the selection and frequency differ by the characteristics of the telemedicine application and the chosen study type. Our systematic review was conducted to close this research gap.

Objectives

This review aims to investigate the following research questions:

1. In which scenarios have PROMs and PREMs been collected for evaluation purposes?
2. Which PROM and PREM outcome domains have been covered and how often?
3. Which OMIs have been used and how often?
4. Did the selection and quantity of PROMs and PREMs differ between study types and application types?
5. Has the use of PROMs and PREMs in evaluation studies changed over time?

Furthermore, we will assess the extent to which the results can be transferred to use cases that have been derived from frequent combinations of application types and medical purposes.

Methods

Systematic Literature Research

To identify relevant articles, we conducted an electronic database search on MEDLINE and Embase. On the basis of the Population, Intervention, Comparison, Outcome, Studies scheme, the following inclusion and exclusion criteria were defined (Textbox 1):

Textbox 1. Inclusion and exclusion criteria.

Patients

- Inclusion criteria
  - All patient groups with an indication for telemedicine care
- Exclusion criteria
  - No patient group using telemedicine

Intervention

- Inclusion criteria
  - Telemedicine applications with patients as main users
- Exclusion criteria
  - Telemedicine applications with no patients as main users, for example, telecommunication between health professionals
  - Telemedicine services containing a single telephone call or electronic message
  - Telemedicine intervention addresses more than one International Statistical Classification of Diseases and Related Health Problems, 10th revision chapter (however, multiple conditions allowed within one International Statistical Classification of Diseases and Related Health Problems, 10th revision chapter); no telemedicine

Control

- Inclusion criteria
  - Nontelemedical standard care (treatment as usual) or prospective designs
- Exclusion criteria
  - Telemedicine versus telemedicine

Outcome

- Inclusion criteria
  - Patient-reported outcome measures or patient-reported experience measures
- Exclusion criteria
  - No patient-reported outcome measures or patient-reported experience measures

Studies

- Inclusion criteria
  - Feasibility studies, noncontrolled trials, controlled trials, and randomized controlled trials
  - Publications in English or German language
  - No limitations on the date of publication
- Exclusion criteria
  - Papers about telemedicine in general, guidelines and handbooks
  - Reviews
  - Case reports
  - Retrospective studies
  - Qualitative studies
  - No English or German language

The search string (Multimedia Appendix 1) was based on 2 previous studies. The part dealing with the assessment of telemedicine applications is based on a review by Arnold and Scheibe et al [4], which aimed to identify standards for the

evaluation of telemedicine applications. The part of the search string covering PROMs and PREMs is based on the PROM Group Construct and Instrument Type Filters of the University of Oxford [17]. This search string has already proven itself in the design of other reviews [18,19]. The search query was performed on April 2, 2020.

Development of Data Extraction Matrix and Used Taxonomies

A matrix was developed as the basis for data extraction. The studies were categorized by (1) study type (feasibility study, noncontrolled trial, controlled trial, and randomized controlled trial [RCT]), (2) medical purpose (first letter of International Statistical Classification of Diseases and Related Health Problems, 10th revision [ICD-10] classification [20]), and (3) application type based on the taxonomy developed by Harst et al [21,22]. This taxonomy was chosen because of its development based on empirical data, which allows its use in quantifying and statistically analyzing the characteristics of telemedicine applications. This taxonomy differentiates between 6 different application types: (1) teleconsultation, a process of providing health care from health care providers to patients over a distance [23]; (2) telediagnostics, a process where a disease is identified over a distance [24]; (3) teleambulance or tele-emergency, a process where emergency care is assisted or data are collected during an emergency over a distance [25]; (4) telemonitoring, a process of data collection over a distance for the purpose of medical decision-making [23,26,27]; (5) telerehabilitation, a process of data collection over a distance for the purpose of coping with the long-term consequences of a disease or an impairment [28]; and (6) digital self-management, a process to promote responsibility for one's own health and to encourage health literacy [29,30]. The classification into application types is intended to be the basis for subsequent subgroup analyses and has already been proven useful for this purpose in other systematic reviews evaluating telemedicine interventions [31,32].

All studies have been reviewed for the use of PROMs and PREMs; both could be represented by established and potentially validated OMIs, which were used frequently in nontelemedicine trials, or OMIs developed especially for the study in question. The OMIs were checked to verify whether they were established instruments or had been developed specifically for a study (SELF_PROM and SELF_PREM). The availability of a validation study served as an indicator of an established instrument. The psychometric properties of the OMIs were irrelevant for the classification into established and self-developed measures, as assessing the quality of the instrument was not within the scope of the review. The assignment of the OMIs to the individual outcome domains took place in an iterative process. In the first step, paraphrases were freely assigned to the OMIs. In the second step, the paraphrases were collected, mapped, and the corresponding categories were developed by the reviewers (AK and SH). The preliminary work of the Core Outcome Measures in Effectiveness Trials initiative provided the framework for the development of categories [33] but was supplemented by additional domains or modified where required. This was necessary, as the Core Outcome Measures in Effectiveness Trials initiative's taxonomy does not sufficiently describe and categorize PREMs to fit the purpose of this review; thus, they had to be developed inductively from the collected and mapped paraphrases. Furthermore, categories were assigned to either the PROM or PREM areas. In the third step, OMIs were assigned to the previously defined outcome domains. To ensure objectivity in the assignment of outcome domains, the reviewers wrote a codebook in advance (Table 1).

Table 1. Codebook of the outcome domains.

Domain	Description[a]
PROM[b]	
HRQoL[c]	Measures the HRQoL of the respondent
Physical function	Measures the extent to which the illness affects the physical function of the respondent
Social function	Measures the extent to which the illness affects the social function of the respondent
Emotional function	Measures the extent to which the illness affects the emotional function of the respondent
Cognitive function	Measures the extent to which the illness affects the cognitive function and disease perception of the respondent
Health literacy	Measures the respondent's ability to avoid, alleviate, or live with a disease
Side effects	Measures complaints caused by therapeutic measures
Adherence	Measures the active role of the patient in the implementation of a therapy
PREM[d]	
Treatment	Deals with the experience of the medical component of a telemedical intervention
Technology	Deals with the experience of the technical component of a telemedical intervention
Satisfaction	Measures the general or overarching satisfaction with the telemedical intervention; satisfaction does not specifically target the medical or technical components of a telemedical intervention

[a]The domain contains outcome measurement instruments.

[b]PROM: patient-reported outcome measure.

[c]HRQoL: health-related quality of life.

[d]PREM: patient-reported experience measure.

Data Extraction

The developed matrix provided the basis for subsequent data extraction. The extraction of paper characteristics and information concerning study type, medical purpose, and application type was performed by 1 reviewer (AK) because of the limited risk of misinterpretation. A total of 2 reviewers (AK and SH) independently performed the assignment of OMIs to PROM and PREM outcome domains based on the developed codebook. In case of any disagreement, assignments were discussed and resolved by consent. The complete data extraction matrix can be found in Multimedia Appendix 1.

Statistical Analysis

For the descriptive analysis, absolute and relative frequencies, mean values, and SDs were calculated for the individual outcome domains and for PROMs and PREMs. The calculations were performed once for all included studies as a whole and also individually for all study and application types. Correlation analyses according to Pearson for metric data and Kendall tau-b for ordinal data were performed to check the strength of dependencies.

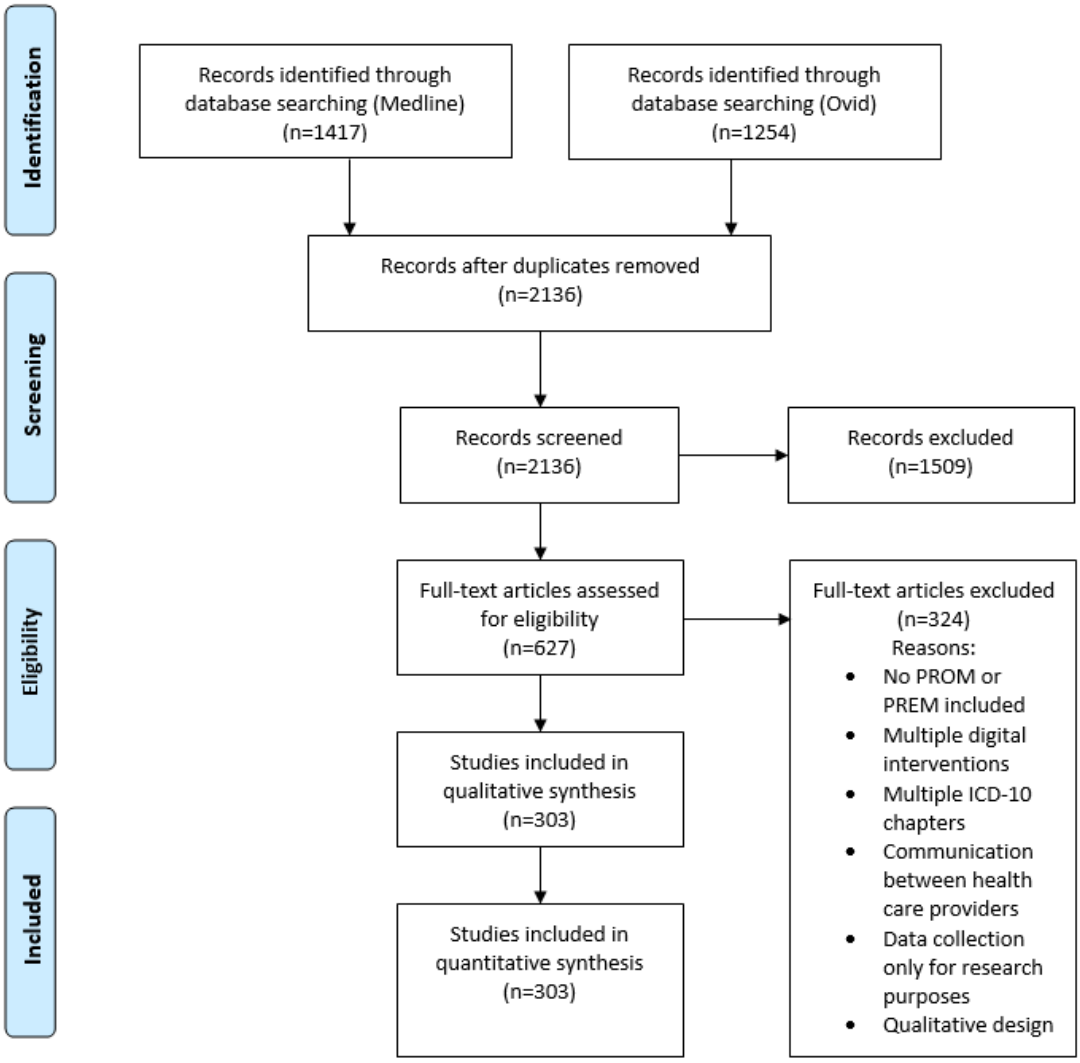
To examine the transfer of results to individual subgroups, 3 use cases were selected from frequent combinations of medical purpose and application types. For this purpose, the frequent outcome domains and study types were determined and descriptively compared with the overall results.

Results

Study Selection

Overall, the electronic search resulted in 2671 hits. Of the 2671 studies, 2136 (79.97%) studies were included in the title abstract screening after removing duplicates. A total of 2 reviewers (AK and LH) performed this step. AK screened all the papers, and LH screened a sample to validate AK's screening. The match between the reviewers was 82.3%, which, according to the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) guidelines [34], legitimizes the examination of only a sample by a second reviewer. Of the 2136 papers, 627 (29.35%) papers were selected for full-text screening, which could be conducted by 1 reviewer (AK) because of the strictly formulated inclusion and exclusion criteria. Of the 627 papers, 303 (48.3%) papers were included in the review (Figure 1). A complete list of all inclusions can be found in Multimedia Appendix 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart. ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th revision; PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



Telemedicine Scenarios

All included studies (n=303) were categorized according to their medical purpose in terms of the ICD-10 chapter and the telemedicine application type (Table 2). The most common ICD-10 chapters were *I* for diseases of the circulatory system (51/303, 16.8%), *C* for neoplasm (47/303, 15.5%), and *F* for mental and behavioral disorders (44/303, 14.5%). Studies that could not clearly be assigned to a chapter were summarized under the term *other* (40/303, 13.2%). These studies were usually telemedicine applications from the fields of *primary prevention*, *aging*, and *well-being*.

Table 2. Identified scenarios of telemedicine applications evaluated via patient-reported outcome measures and patient-reported experience measures.

Application type	Teleambulance (N=0), n	Telediagnostics (N=4), n	Digital self-management (N=78), n	Teleconsultation (N=75), n	Telemonitoring (N=96), n	Telerehabilitation (N=50), n
ICD-10[a] chapter						
A[b] (N=1), n	0	0	1	0	0	0
B[b] (N=9), n	0	0	4	3	2	0
C[c] (N=47), n	0	0	11	11	21	4
D[c,d] (N=0), n	0	0	0	0	0	0
E[e] (N=24), n	0	0	9	10	4	1
F[f] (N=44), n	0	1	13	22	6	2
G[g] (N=15), n	0	0	3	5	4	3
H[h] (N=3), n	0	1	0	2	0	0
I[i] (N=51), n	0	1	5	2	22	21
J[j] (N=19), n	0	0	3	2	9	5
K[k] (N=12), n	0	0	8	0	4	0
L[l] (N=6), n	0	1	3	3	1	0
M[m] (N=17), n	0	0	2	1	7	7
N[n] (N=6), n	0	0	2	3	1	0
O[o] (N=1), n	0	0	0	0	0	1
P[p] (N=0), n	0	0	0	0	0	0
Q[q] (N=2), n	0	0	1	0	1	0
R[r] (N=0), n	0	0	0	0	0	0
S[s] (N=2), n	0	0	0	0	0	2
T[t] (N=2), n	0	0	0	0	1	1
V[u] (N=0), n	0	0	0	0	0	0
Z[v] (N=0), n	0	0	0	0	0	0
Other (N=40), n	0	0	14	10	13	3

[a]ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th revision.

[b]A-B: certain infectious and parasitic diseases.

[c]C-D: neoplasms.

[d]D: diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism.

[e]E: endocrine, nutritional, and metabolic diseases.

[f]F: mental, behavioral, and neurodevelopmental disorders.

[g]G: diseases of the nervous system.

[h]H: diseases of the eye and adnexa; diseases of the ear and mastoid process.

[i]I: diseases of the circulatory system.

[j]J: diseases of the respiratory system.

[k]K: diseases of the digestive system.

[l]L: diseases of the skin and subcutaneous tissue.

[m]M: diseases of the musculoskeletal system and connective tissue.

[n]N: diseases of the genitourinary system.

[o]O: pregnancy, childbirth, and the puerperium.

[p]P: certain conditions originating in the perinatal period.

[q]Q: congenital malformations, deformations, and chromosomal abnormalities.

[r]R: symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified.

[s]S: injury, poisoning, and certain other consequences of external causes.

[t]T: injury, poisoning, and certain other consequences of external causes.

[u]V: external causes of morbidity.

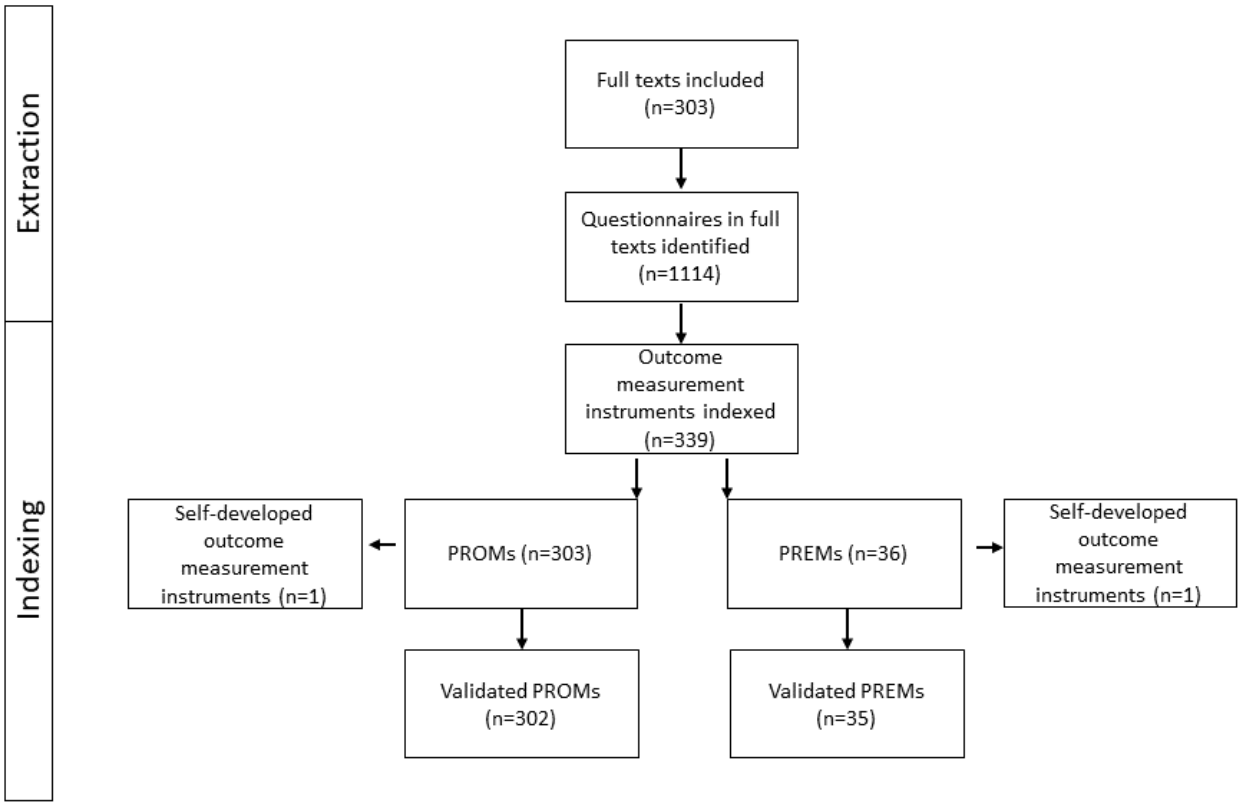
[v]Z: factors influencing health status and contact with health services.

Telemonitoring (96/303, 31.7%) was the most frequent type of application, followed by digital self-management (79/303, 26.1%), teleconsultation (75/303, 24.8%), and telerehabilitation (50/303, 16.5%), telediagnostics (4/303, 1.3%); there were no studies with teleambulance (0/303, 0%). The most common combinations of medical purpose and application type were diseases of the circulatory system+telemonitoring (22/303, 7.3%), mental and behavioral disorders+teleconsultation (22/303, 7.3%), diseases of the circulatory system+telerehabilitation (21/303, 6.9%), and neoplasm+telemonitoring (21/303, 6.9%). All other combinations were found in <20 cases. Of the 144 possible combinations, only 51 (35.4%) were identified in this study.

Use of Outcome Domains

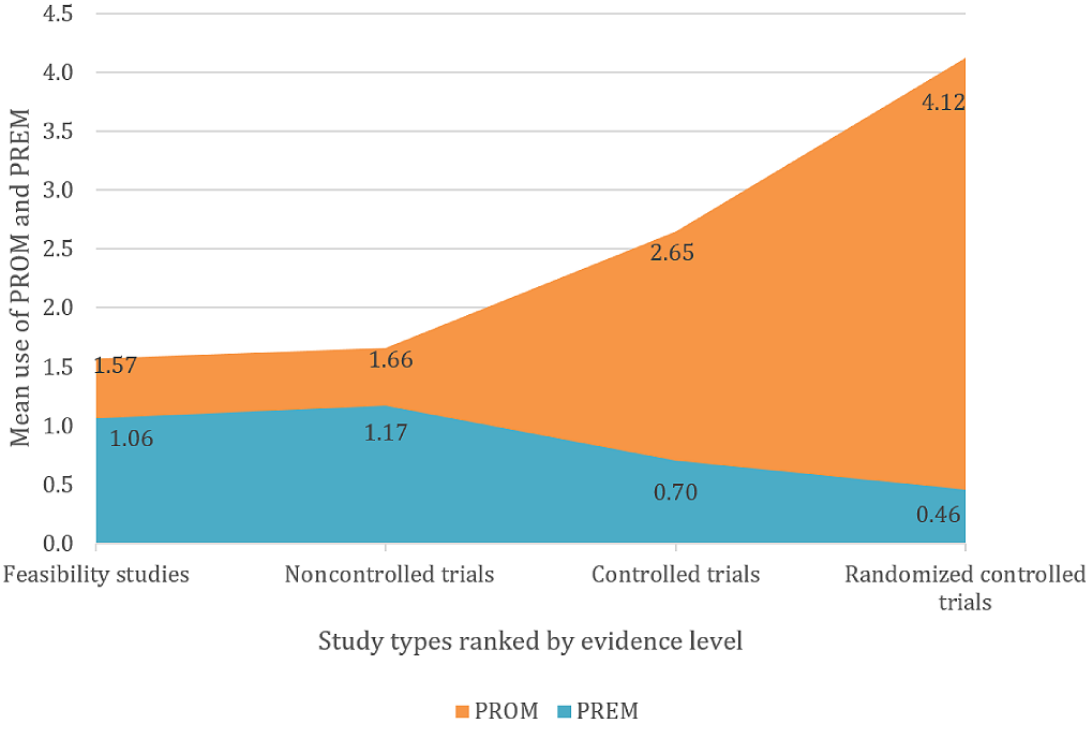
In total, 339 different OMIs were used in 1114 cases in the included studies (n=303). The OMIs were classified into 89.4% (303/339) PROMs and 10.6% (36/339) PREMs (Figure 2). Measurement instruments, which were developed especially for the individual study and were not listed in databases for PROMs and PREMs, were summarized in SELF_PROM or SELF_PREM. Measurement instruments for general satisfaction with the entire medical treatment process were summed up under the term SAT for satisfaction, which belongs to the field of PREMs and includes various forms of Likert scales, visual analog scales, and other self-developed constructs.

Figure 2. Extraction process of outcome measurement instruments. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



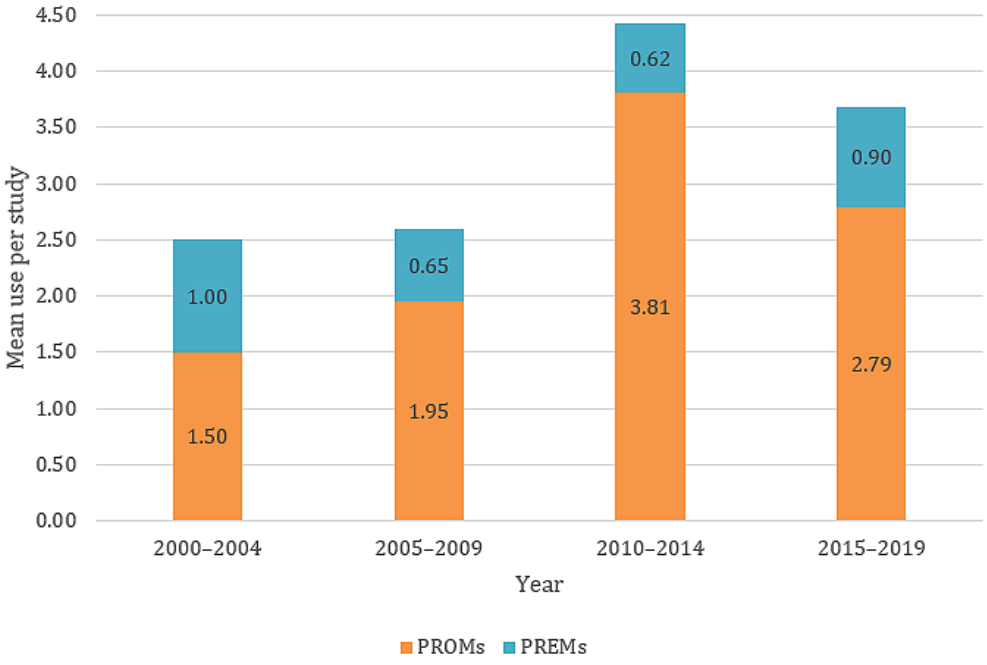
Considering all studies, PROMs (881/1114, 79.08%) were used more frequently than PREMs (233/1114, 20.92%). The correlation analysis indicated that with an increasing number of PROMs, the number of PREMs decreased (r=−0.23; Figure 3). Across all studies, 21.4% (64/303) of PROMs and 22.3% (68/303) of PREMs were self-developed. The frequency of PROMs used was as follows (in descending order): HRQoL (310/881, 35.2%), emotional function (244/881, 27.7%), adherence (103/881, 11.7%), SELF_PROM (77/881, 8.7%), physical function (57/881, 6.5%), cognitive function (38/881, 4.3%), health literacy (35/881, 4%), social function (9/881, 1%), and side effects (8/881, 0.9%). The frequency of PREMs used was as follows (in descending order): general satisfaction (98/233, 42.1%), SELF_PREM (84/233, 36.1%), treatment (29/233, 12.4%), and technology (22/233, 9.4%).

Figure 3. Use of patient-reported outcome measures and patient-reported experience measures by study type. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



Considering the number of collected OMIs per study, it became apparent that most studies used 2-3 OMIs. The maximum number of OMIs used per study was 13 (Figure 4). Most OMIs used were PROMs (used in 881/1114, 79.08% of the included studies). In 15.5% (47/303) of the studies, no PROMs were used. The maximum was 11 PROMs per study (3/303, 1%). No PREMs were collected in 45.9% (139/303) of the studies. In 38.6% (117/303) of studies, one PREM was collected per study. The number declined sharply to 10.6% (32/303) of studies in which 2 PREMs were collected and fell further to the maximum of 0.3% (1/303) of studies in which 6 PREMs were collected.

Figure 4. Mean use of patient-reported outcome measures and patient-reported experience measures over time. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



Outcome Measurement Instruments

The most commonly used PROM OMIs were the HRQoL OMIs EuroQol five-dimension scale [35] in 14.5% (44/303) of studies, the Short Form 36 [36] in 11.9% (36/303) of studies, and emotional function, especially depression symptoms, measured by the Patient Health Questionnaire-9 [37] in 8.9% (27/303) of studies.

The PREM OMIs that were most commonly used were the Client Satisfaction Questionnaire-8 to measure treatment satisfaction [38] in 4% (12/303) of studies and the System Usability Scale usability OMIs in the domain technology [39] in 2% (6/303) of studies. The third most frequently used OMI was the Patient Assessment of Chronic Illness Care OMI, which also measures treatment satisfaction [40], in 1% (3/303) of studies, together with the Telehealth Acceptance Measure [41], used in 1% (3/303) of studies.

The 3 most frequently used OMIs per outcome domain are listed in Table 3.

Table 3. Most frequently used outcome measurement instrument per outcome domain.

Outcome measurement instrument	Studies in the domain, N	Absolute frequency (n) and share in the domain, n (%)	Share in all studies (N=303), n (%)
PROM[a]			
HRQoL[b]			
EuroQol five-dimension scale	307	44 (14.3)	44 (14.5)
Short Form 36	307	36 (11.7)	36 (11.9)
Short Form 12	307	19 (6.2)	19 (6.3)
Physical function			
International Physical Activity Questionnaire	57	7 (12.3)	7 (0.2)
Nottingham Extended Activities of Daily Living Scale	57	4 (7)	4 (1.3)
Active Australia Survey, Activities-specific Balance Scale, and Physical Activity Scale for the Elderly	57	3 (5.2)	3 (1)
Social function			
Work Productivity and Activity Impairment Questionnaire	9	2 (22.2)	2 (0.7)
CHAMPS Activities Questionnaire for Older Adults, World Health Organization Health and Work Performance Questionnaire, Social Phobia Screening Questionnaire, and others	9	1 (11.1)	1 (0.3)
Emotional function			
Patient Health Questionnaire-9	244	27 (11.1)	27 (8.9)
Hospital Anxiety and Depression Scale	244	23 (9.4)	23 (7.6)
Center for Epidemiologic Studies Depression Scale	244	16 (6.6)	16 (5.3)
Cognitive function			
Brief Illness Perception Questionnaire	41	3 (7.3)	3 (1)
Supportive Care Needs Survey Short Form 34, Supportive Care Needs Survey Screening Tool 9, and Illness Perception Questionnaire	41	2 (4.9)	2 (0.7)
Body Attitude Test, Functional Activities Questionnaire, Illness Cognition Questionnaire, and others	41	1 (2.4)	1 (0.3)
Health literacy			
European Heart Failure Self-Care Behaviour Scale and Self-Care of Heart Failure Index	35	3 (8.6)	3 (1)
Health Education Impact Questionnaire, Health Promoting Lifestyle Profile II, Patient Enablement Instrument, and others	35	2 (5.7)	2 (0.7)
Cancer Empowerment Questionnaire, Diabetes Self-Management Profile, Revised Heart Failure Compliance, and others	35	1 (2.9)	1 (0.3)
Side effects			
Patient Neurotoxicity Questionnaire, Glasgow Antipsychotic Side-Effect Scale, Side Effects of Anti-epileptic Drugs, and others	8	1 (12.5)	1 (0.3)
Adherence			
Morisky Medication Adherence Scale	103	5 (4.9)	5 (1.7)
Medication Adherence Rating Scale	103	4 (3.9)	4 (1.3)
AIDS Clinical Trails Group Adherence Questionnaire	103	1 (1)	1 (0.3)
PREM[c]			
Treatment			
Client Satisfaction Questionnaire	29	12 (41.4)	12 (4)
Diabetes Treatment Satisfaction Questionnaire, Patient Assessment of Chronic Illness Care, and Patient Satisfaction Questionnaire Short Form	29	2 (6.9)	(0.7)

Outcome measurement instrument	Studies in the domain, N	Absolute frequency (n) and share in the domain, n (%)	Share in all studies (N=303), n (%)
Canadian Health Care Evaluation Project questionnaire, Patient Experience Questionnaire, Functional Assessment of Chronic Illness Therapy–Treatment Satisfaction–Patient Satisfaction, and others	29	1 (3.4)	1 (0.3)
Technology			
System Usability Scale	22	6 (27.3)	6 (2)
Telehealth Acceptance Measure	22	3 (13.6)	3 (1)
Post-Study System Usability Questionnaire and Usefulness, Satisfaction, and Ease of use Questionnaire	22	2 (12.3)	2 (0.7)

[a]PROM: patient-reported outcome measure.

[b]HRQoL: health-related quality of life.

[c]PREM: patient-reported experience measure.

On average, each OMI was used 3.29 times, compared across all studies; however, most OMIs were only used once (modal value=1). There was a large variation in the frequency of use (SD 8.45) of single OMIs. Considering the frequency of use of single OMIs within the respective outcome domains, even the most frequently used OMIs, only achieved shares of ≤20% in the respective domains in most cases. This indicates a high heterogeneity of PROMs and PREMs used in the single outcome domains. To show this in a more differentiated manner, Table 4 indicates the absolute number of non–self-developed OMIs per outcome domain and their absolute frequency of use.

Table 4. Outcome measurement instrument per outcome domain in absolute numbers.

Outcome domain	Outcome measurement instruments (N=337), n (%)	Absolute frequency of uses (N=953), n (%)
PROM[a]		
HRQoL[b]	109 (32.3)	310 (32.5)
Physical function	35 (10.4)	57 (6)
Social function	8 (2.4)	9 (0.9)
Emotional function	92 (27.3)	244 (25.6)
Cognitive function	28 (8.3)	38 (4)
Health literacy	16 (4.7)	35 (3.7)
Side effects	8 (2.4)	8 (0.8)
Adherence	6 (1.8)	103 (10.8)
PREM[c]		
Treatment	14 (4.2)	29 (3)
Technology	14 (4.2)	22 (2.3)

[a]PROM: patient-reported outcome measure.

[b]HRQoL: health-related quality of life.

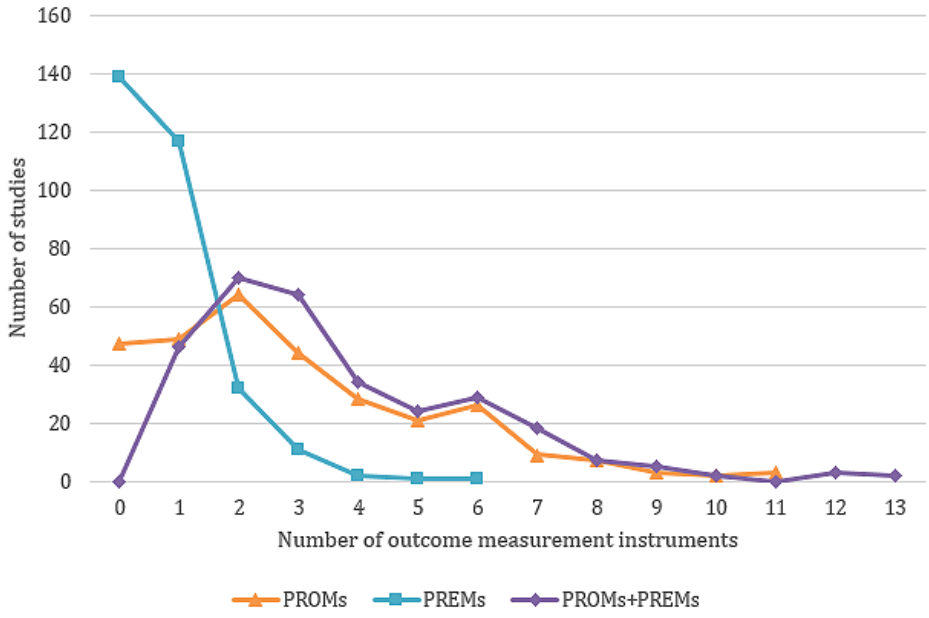
[c]PREM: patient-reported experience measure.

OMIs that were developed explicitly for use in telemedicine applications could only be identified for PREMs. These 6 OMIs were the Telehealth Acceptance Measure (3/233, 1.3%) [41], Mobile Application Rating Scale (1/233, 0.4%) [42], Patient Assessment of Communication during Telehealth (1/233, 0.4%) [43], Service User Technology Acceptability Questionnaire (1/233, 0.4%) [44], Telemedicine Perception Questionnaire (1/233, 0.4%) [45], and Telehealth Usability Questionnaire (1/233, 0.4%) [46]. Telemedicine-specific questionnaires were used in only 3.4% (8/233) of all PREMs.

Chronological Trends in the Use of PROMs and PREMs

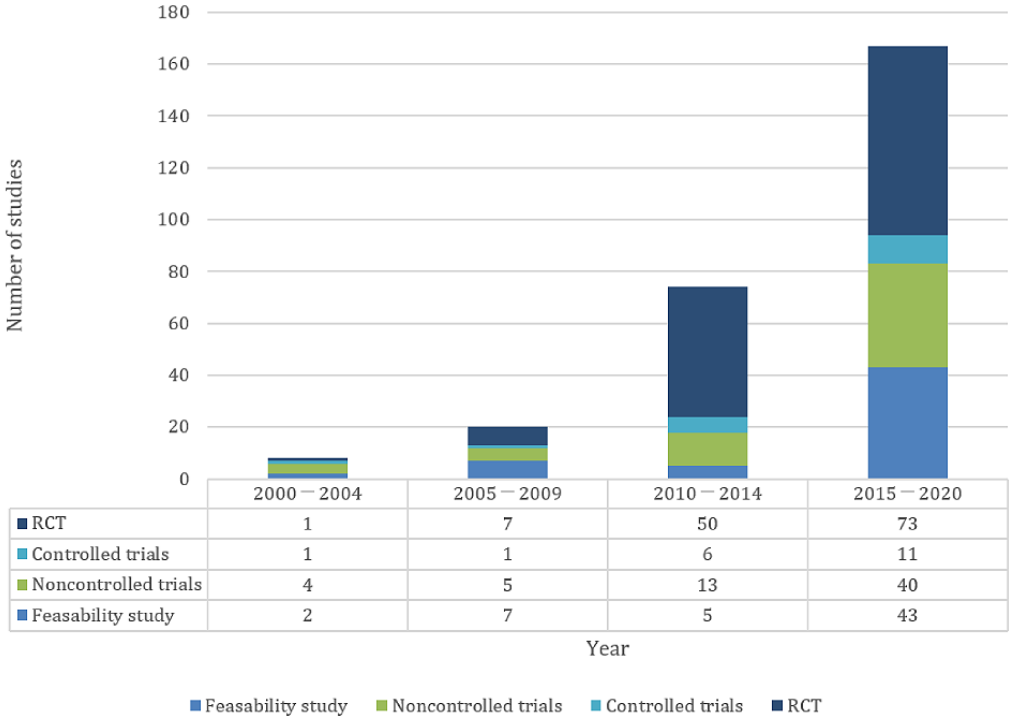
The included studies were clustered into 5-year groups for analysis of the evaluation practice development over time (Figure 5). The year 2020 was not included in the analysis, as data were only available for the first 4 months of that year. The number of included studies increased above average over the years. The share of RCTs doubled every 5 years until 2014 and then dropped from 68.5% (50/73) to 43.7% (73/166) from 2014 to 2019.

Figure 5. Number of outcome measurement instruments collected per study. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



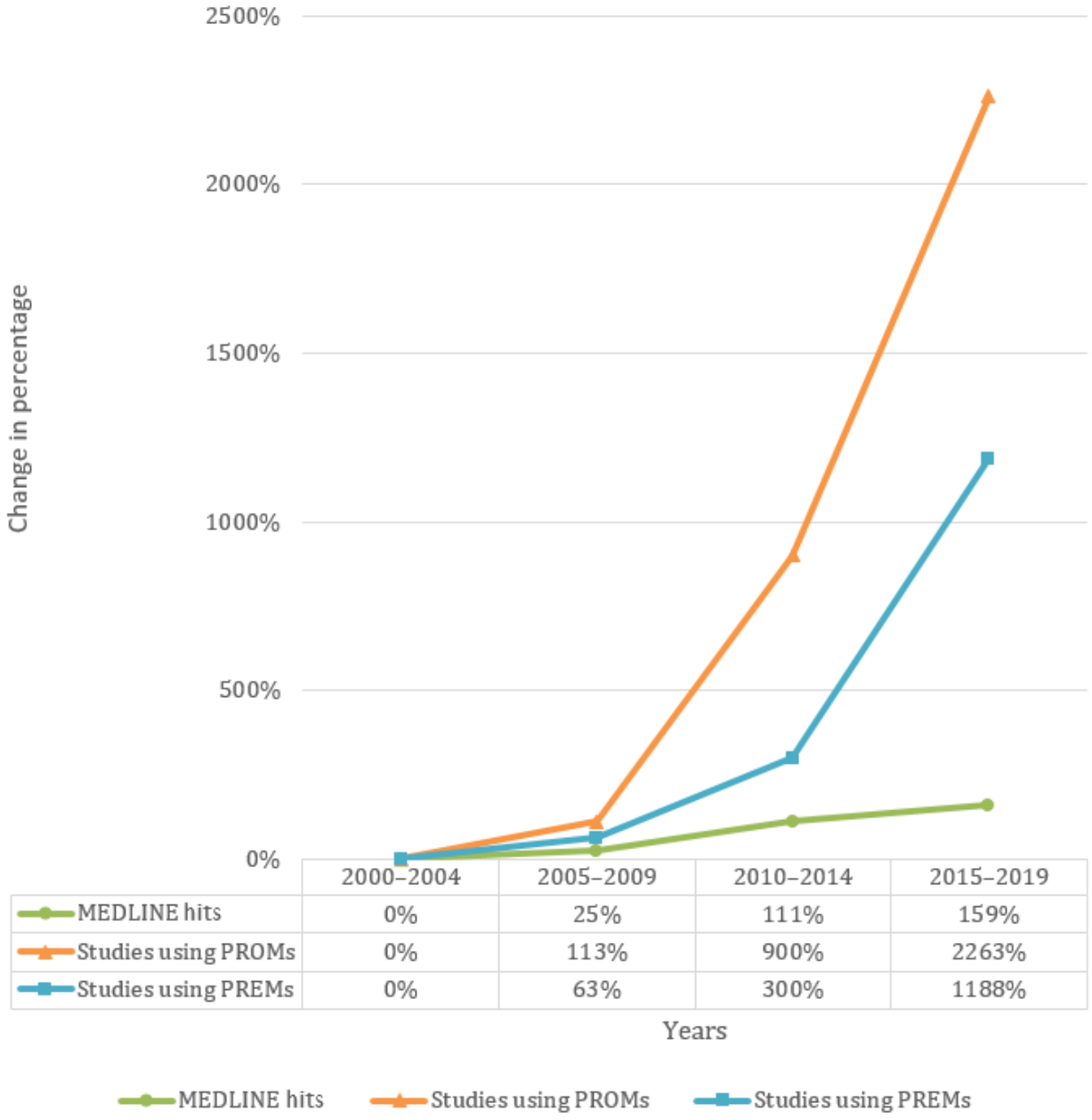
The average use of PROMs per study, as well as the total number of OMIs used, steadily increased between 2000 and 2014 and then decreased between 2015 and 2019 (Figure 6). The mean use of PREMs per study remained at a lower level permanently compared with PROMs.

Figure 6. Numbers of studies by study type over time. RCT: randomized controlled trial.



To examine the change in the evaluation of telemedicine over time, the 2000-2004 episode was used as a starting point (Figure 7). The percentage increase or decrease compared with that in 2000-2004 was calculated. In addition, the number of telemedicine studies, regardless of whether they used a single PROM and PREM, was determined by hits of the term *telemedicine* in MEDLINE per year. These were compared with the included studies that used PROMs and PREMs for evaluation.

Figure 7. Change over time. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



The number of telemedicine studies has steadily increased over time. However, the number of studies reporting PROMs and the number of studies reporting PREMs increased more compared with MEDLINE hits.

Subgroup Analysis: Application Type

Subgroup analysis for application type was conducted to cluster the technologies described in the studies according to their intended medical purpose and to explore differences in the evaluation approaches. On average, more PROMs were applied in studies focusing on telerehabilitation (mean 3.82, SD 2.60) and digital self-management (mean 3.51, SD 2.51) than on teleconsultation (mean 2.63, SD 2.41), telemonitoring (mean 2.24, SD 1.92), and telediagnostics (mean 1.00, SD 2.00). The application of PREMs was distributed evenly across all application types (range of mean values 0.50-1.06). Figure 8 shows the mean values of the PROMs and PREMs used by application type and compared with the mean values of all studies. The values for all the application types and outcome domains can be found in Table 5.

Figure 8. Use of patient-reported outcome measures and patient-reported experience measures by application type. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.

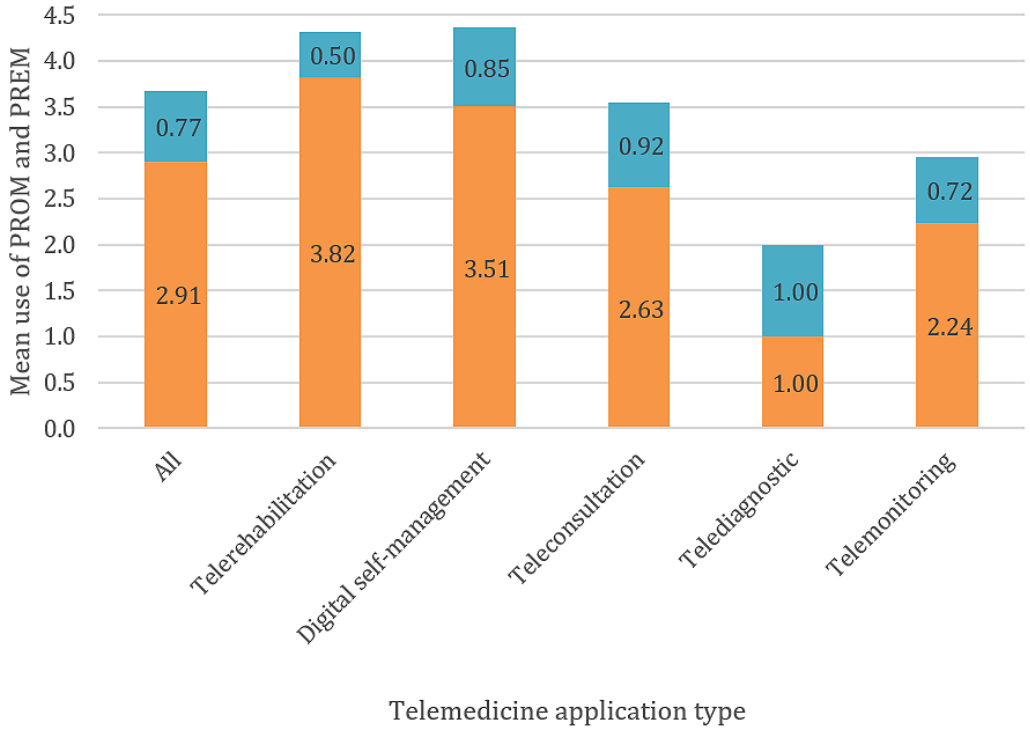


Table 5. Outcomes by application type (N=303).

Outcomes	All		Telediagnostics (n=4)		Digital self-management (n=78)		Teleconsultation (n=75)		Telemonitoring (n=96)		Telerehabilitation (n=50)	
	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)
PROM[a]												
PROM (total)	884	2.91 (2.40)	4	1.00 (2.00)	274	3.51 (2.49)	197	2.63 (2.41)	215	2.24 (1.92)	191	1.40 (2.60)
HRQoL[b]	310	1.02 (1.05)	1	0.25 (0.50)	78	1.00 (0.50)	65	0.87 (1.00)	94	0.98 (0.99)	72	1.44 (1.05)
Physical function	57	0.19 (0.45)	0	0.00 (0.00)	17	0.22 (0.47)	6	0.08 (0.27)	8	0.08 (0.28)	26	0.52 (0.71)
Social function	9	0.03 (0.17)	0	0.00 (0.00)	5	0.06 (0.25)	3	0.04 (0.20)	0	0.00 (0.00)	1	0.02 (0.14)
Emotional function	244	0.80 (1.18)	2	0.50 (1.00)	82	1.05 (1.32)	75	1.00 (1.48)	42	0.44 (0.81)	43	0.86 (0.90)
Cognitive function	38	0.13 (0.40)	0	0.00 (0.00)	13	0.17 (0.37)	9	0.12 (0.37)	6	0.06 (0.28)	10	0.20 (0.61)
Health literacy	37	0.12 (0.38)	1	0.25 (0.50)	12	0.15 (0.45)	4	0.05 (0.28)	14	0.15 (0.43)	4	0.06 (0.27)
Side effects	8	0.03 (0.18)	0	0.00 (0.00)	4	0.05 (0.27)	0	0.00 (0.00)	2	0.02 (0.14)	2	0.04 (0.20)
Adherence	103	0.34 (0.53)	0	0.00 (0.00)	35	0.45 (0.57)	19	0.25 (0.44)	30	0.31 (0.55)	19	0.38 (0.53)
Self_PROM	77	0.25 (0.52)	0	0.00 (0.00)	28	0.36 (0.60)	16	0.21 (0.47)	19	0.20 (0.45)	14	0.28 (0.57)
PREM[c]												
PREM (total)	234	0.77 (0.92)	4	1.00 (0.00)	66	0.85 (0.93)	69	0.92 (1.06)	69	0.72 (0.85)	25	0.50 (0.79)
Treatment	29	0.10 (0.32)	0	0.00 (0.00)	10	0.13 (0.37)	9	0.12 (0.37)	7	0.07 (0.26)	3	0.06 (0.24)
Technology	23	0.08 (0.31)	1	0.25 (0.50)	7	0.09 (0.30)	1	0.01 (0.12)	7	0.07 (0.26)	6	0.12 (0.52)
Satisfaction	98	0.32 (0.53)	2	0.50 (0.58)	26	0.33 (0.55)	35	0.47 (0.60)	26	0.27 (0.49)	9	0.18 (0.39)
Self_PREM	84	0.28 (0.59)	1	0.25 (0.50)	23	0.29 (0.58)	24	0.32 (0.64)	29	0.30 (0.63)	7	0.14 (0.4)

[a]PROM: patient-reported outcome measure.

[b]HRQoL: health-related quality of life.

[c]PREM: patient-reported experience measure.

Subgroup Analysis: Study Type

The second subgroup analysis was conducted based on the study type to evaluate the use frequency of PROMs and PREMs in different types of studies and the levels of evidence they were associated with. Of the 303 studies, 67 (22.1%) feasibility studies, 70 (23.1%) noncontrolled trials, 20 (6.6%) controlled trials, and 146 (48.2%) RCTs were identified. The study design served as an indicator of the evidence level of the studies [5]. The evidence level was determined according to the guidelines of the Oxford Centre for Evidence-based Medicine [47]. Study types with evidence level 3, such as feasibility studies (mean 1.66, SD 1.66) and noncontrolled trials (mean 1.66, SD 1.64), used fewer PROMs than controlled trials (mean 2.65, SD 2.72), with evidence level 2 or even RCTs (mean 4.12, SD 2.36), with evidence level 1. An opposite trend was observed for PREMs. The values for PREMs in order of increasing evidence level were as follows: feasibility study (mean 1.22, SD 0.87), noncontrolled trial (mean 1.00, SD 1.14), controlled trial (mean 0.70, SD 0.86), and RCT (mean 0.46, SD 0.71). The correlation analysis for the relationship between the number of PROMs or PREMs and the evidence levels resulted in r=−0.50 for PROMs and r=0.34 for PREMs (Figure 3). Table 6 lists the complete distribution of outcomes by study type.

Table 6. Outcomes by study type.

Outcomes	All (n=301)		Feasibility study (n=67)		Noncontrolled trial (n=70)		Controlled trial (n=20)		Randomized controlled trial (n=146)	
	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)	Σ	Mean value (SD)
PROM[a]										
PROM	884	2.91 (2.40)	111	1.66 (1.66)	116	1.66 (1.64)	53	2.65 (2.72)	601	4.12 (2.36)
HRQoL[b]	310	1.02 (1.05)	33	0.49 (0.79)	38	0.54 (0.79)	22	1.10 (0.91)	217	1.49 (1.07)
Physical function	57	0.19 (0.45)	3	0.04 (0.21)	4	0.30 (0.23)	3	0.15 (0.49)	47	0.32 (0.56)
Social function	9	0.03 (0.17)	0	0 (0)	2	0.03 (0.17)	0	0 (0)	7	0.05 (0.21)
Emotional function	244	0.80 (1.18)	25	0.37 (0.69)	32	0.00 (0.90)	11	0.55 (1.00)	176	1.21 (1.36)
Cognitive function	38	0.13 (0.40)	4	0.06 (0.24)	4	0.46 (0.23)	6	0.30 (0.57)	24	0.16 (0.57)
Health literacy	37	0.12 (0.38)	8	0.12 (0.41)	5	0.06 (0.26)	3	0.15 (0.49)	19	0.13 (0.38)
Side effects	8	0.03 (0.18)	3	0.04 (0.27)	1	0.07 (0.12)	0	0 (0)	4	0.03 (0.16)
Adherence	103	0.34 (0.53)	21	0.31 (0.53)	15	0.01 (0.41)	3	0.15 (0.37)	64	0.44 (0.58)
Self_PROM	77	0.25 (0.52)	14	0.21 (0.41)	15	0.21 (0.45)	5	0.25 (0.44)	43	0.29 (0.60)
PREM[c]										
PREM	234	0.77 (0.92)	82	1.22 (0.87)	70	1.00 (1.44)	14	0.70 (0.86)	67	0.46 (0.71)
Treatment	29	0.10 (0.32)	4	0.06 (0.24)	3	0.04 (0.20)	1	0.05 (0.22)	21	0.14 (0.39)
Technology	23	0.08 (0.31)	12	0.18 (0.49)	3	0.04 (0.20)	0	0 (0)	7	0.05 (0.24)
Satisfaction	98	0.32 (0.53)	32	0.48 (0.59)	34	0.49 (0.63)	7	0.35 (0.59)	25	0.17 (0.38)
Self_PREM	84	0.28 (0.59)	34	0.51 (0.68)	30	0.43 (0.77)	6	0.30 (0.57)	14	0.10 (0.34)

[a]PROM: patient-reported outcome measure.

[b]HRQoL: health-related quality of life.

[c]PREM: patient-reported experience measure.

Use Cases

Three use cases were formed to check the results for transferability and were based on common combinations of medical purpose and application type. The use cases were telemonitoring for cancer diseases (21/303, 6.9%), teleconsultation for mental and behavioral disorders (22/303, 7.3%), and telerehabilitation for cardiovascular diseases (21/303, 6.9%). Although the total number of studies on telemonitoring for diseases of the circulatory system was 22, we chose to cover the widest possible range of characteristics within the presented use cases. Therefore, we opted for telemonitoring for cancer diseases and telerehabilitation for cardiovascular diseases, although these have lower numbers.

A descriptive analysis of the distribution of PROMs and PREMs and their outcome domains was also conducted. Again, the ratio of PROMs was different from that of PREMs (Figure 9). Similarly, the proportion of PREMs in the use case of telemonitoring for cancer diseases with evidence level 3 was higher than in the other 2 use cases with evidence level 1. HRQoL and emotional function were found to be the most frequently used outcome domains in all 3 cases (Table 7). Only the third most frequent outcome, satisfaction, was case-specific; it accounted for half of the cases. The results of the entire sample could be transferred to the 3 use cases, which could be an indication of the transferability of the review results to specific use cases.

Figure 9. Use of patient-reported outcome measures and patient-reported experience measures by use cases. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.

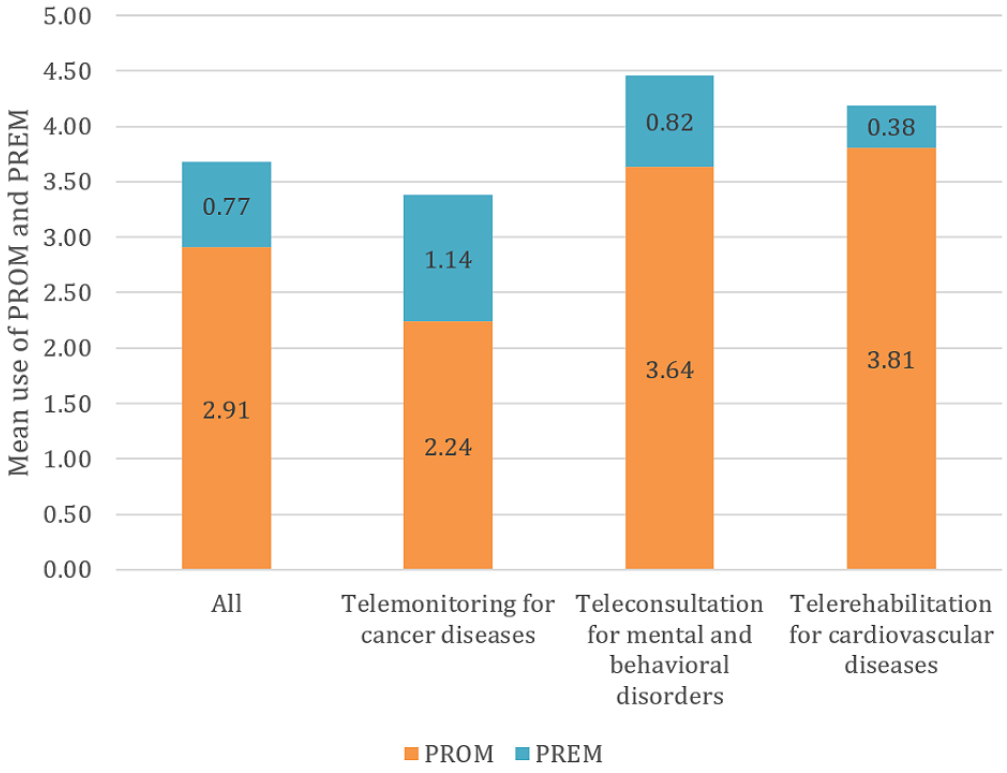


Table 7. Use cases.

Characteristics	Use cases			
	All studies (n=303)	Telemonitoring for cancer diseases (n=21)	Teleconsultation for mental and behavioral disorders (n=22)	Telerehabilitation for cardiovascular diseases (n=21)
Most common outcomes				
1	HRQoL[a]	HRQoL	Emotional function	HRQoL
2	Emotional function	Emotional function	HRQoL	Emotional function
3	Adherence	Satisfaction	Satisfaction and adherence	Physical function
Evidence level				
Modus	1	3	1	1

[a]HRQoL: health-related quality of life.

Discussion

Summary and Discussion of Main Findings

The aim of this systematic review was to empirically examine the characteristics of PROM and PREM use in evaluation studies of telemedicine applications. Owing to the large number of possible combinations of application types (n=6) and medical purposes (n=24), there was great heterogeneity in the evaluation studies. Of the 144 possible combinations, 51 (35.4%) were identified in this study. However, we were able to answer the research questions.

PROMs dominated the evaluation of telemedicine applications. In total, 80% (4/5) of OMIs were PROMs, and only in 14% (1/7) of studies was no PROM used. On the other hand, PREMs were used in less than half of the studies, and hardly any of these PREMs were adapted to telemedical care. The lack of telemedicine-specific OMIs was apparently compensated for by the use of self-developed OMIs. This could indicate that the existing OMIs could not be applied because of the great heterogeneity of the telemedicine-specific use cases, did not collect the desired outcomes, or were simply not known to the evaluation team. The review by Hajesmaeel-Gohari and Bahaadinbeigy [48] in 2021 examined the use of validated telemedicine-specific OMIs in the form of PREMs for the evaluation of telemedicine service quality. The review was able to identify 59 different PREMs, of which only the 10 most frequent were mentioned. Our review was able to identify 70% (7/10) of the most frequent PREMs. However, the frequency distributions of the OMIs used do not match between the two reviews, as Hajesmaeel-Gohari and Bahaadinbeigy [48] identified a higher number of PREMs because of a more specific

search strategy. They concluded that the use of PREMs for the evaluation of the quality of telemedicine applications should be obligatory and needs to be expanded, which also requires the development of further specific OMIs [48].

The quantity of PREMs decreased with an increasing number of PROMs; that is, a negative correlation (r=−0.23) was observed. One explanation for this correlation could be that the number of OMIs and outcome domains was kept as low as possible. In the sample, the median was 3 OMIs and outcome domains per study. However, the number of outcome domains per study varied (SD 2.36). As the OMIs are constructs of several items, depending on the instrument, this can range from a handful to several dozen items; the total number of items should be taken into account when selecting the OMIs [48]. Furthermore, the study participants or patients should not be overwhelmed by the total number of OMIs and included items as this could lead to incomplete answers or even dropout [49].

The number of telemedicine studies that collected PROMs and PREMs increased on average over time (Figures 5 and 6). In addition, the proportion of high-evidence studies, especially RCTs, also increased (Figure 6). It was shown that in years with a high proportion of high-evidence studies, the ratio of PROMs was considerably higher than the ratio of PREMs, as described above. This could be caused by the wider recognition and implementation of PROMs and PREMs [50,51], as can be seen in Figure 7, where the growth rate of studies using PROMs and PREMs is far higher than the growth rate of telemedicine papers in MEDLINE. The trend toward the increased use of PROMs and PREMs is also evident in several medical disciplines, such as oncology [52] and orthopedics [53], as well as in studies for regulatory purposes for medical devices [54].

In addition, guidelines that recommend the use of PROMs and PREMs published in recent years (eg, MAST 2012 [2] and NICE framework 2019 [3]) could have promoted the increased use of PROMs and PREMs over the years. These guidelines also recommend the use of high-evidence study designs. Again, an increased use of RCTs has been noticed since the publication of these guidelines.

Regardless of the telemedicine evaluation tools used, variations can be found between countries regarding the state of PROM and PREM implementation, types of data use, conditions and therapeutic areas, and challenges and success factors for PREM and PROM use [55]. Hence, regional and cultural aspects must be taken into account when developing, translating, and implementing PROMs, especially if they are measured using electronic tools [56]. Furthermore, these aspects have to be considered when evaluating PROM and PREM scores and comparing them between different countries.

The ratio of PROMs to PREMs also depended on the study type and evidence level. Although in low-evidence studies the frequency of PREMs was almost equal to the frequency of PROMs, it decreased with increasing evidence level. At the same time, more outcomes were recorded at high evidence levels (Figure 6). This could be related to the development cycle of telemedicine technologies [5]. Using evidence level as a surrogate parameter for the maturity stage of the application, feasibility studies and proof-of-concept studies increasingly require information on the usability and acceptance of the technology in addition to the clinical effectiveness. On the other hand, PREMs played almost no role in clinical trials with high evidence levels. PROMs clearly dominated in RCTs in relative and absolute numbers. This is also reflected in the Khoja-Durrani-Scott framework for eHealth evaluation [6]. Khoja et al [6] subdivided the development cycle of an eHealth application into 4 phases. The framework recommends focusing on typical PREM domains, such as usability, user-friendliness, and acceptance in the early phases of development. In later phases, evaluation should focus on health outcomes, such as quality of life and health impact, although these should also be recorded in the early phases. The design and evaluation framework for digital health interventions by Kowatsch et al [5] goes one step further and specifies the outcomes as well as the required study designs for each phase. With each phase, the evidence level of the study designs increases, and the focus of the outcomes change according to the needs. The first phase, the preparation phase, includes feasibility and acceptability studies to determine the ease of use and adherence. In the optimization phase, the first evidence of effectiveness, expected benefits, and satisfaction with the quality of the application should be measured. In the later phases, that is, the evaluation and implementation phases, the success of the implementation of digital health applications should be monitored. The fact that the selection of the evaluation design and outcomes should be made according to the stage of development and should have an appropriate level of evidence has also been pointed out by the MAST model [2] and the evaluation principles of Arnold and Scheibe et al [4]. The correlation of the PREMs (τ=0.35) and the PROMs (τ=−0.45) with evidence level indicates that evaluation was performed as described in the guidelines for maturity stage–based evaluation.

A key milestone in the implementation of PROMs and PREMs in evaluation studies of telemedicine interventions was set by Germany in 2020 with the *Digital Care Act*. One significant innovation is that the costs for the use of so-called *digital health applications* will be reimbursed by statutory health insurance [7,57]. As a result, since October 2020, around 90% of the population is entitled to a wide range of mobile health applications in the areas of telerehabilitation, telemonitoring, and digital self-management [57]. Another significant innovation is that the assessment of bankability does not exclusively depend on the medical benefits, which, among clinical and epidemiological outcomes, could be assessed by PROMs, such as HRQoL, but also on the so-called patient-relevant improvement of structure and processes, which are mainly assessed by PROMs and PREMs. Examples of patient-relevant improvement of structure and processes are coping with difficulties in everyday life because of illness, facilitating access to care, health literacy, patient autonomy, reduction of therapy-related expenses, and burdens for patients and their relatives [7]. Medical benefits and patient-relevant improvements of structure and processes are now of equal importance in the approval process of digital health applications, and only one of the outcomes has to be more effective than standard care [7]. This represents a significant increase in the importance of PROMs and PREMs in evaluation studies of telehealth applications. The reason for including patient-relevant

improvement of structure and processes as an outcome in evaluation studies was that digital health applications are considered to improve patient self-efficacy [58] and health-related behaviors, such as adherence [59] and health literacy [60]. In our review, 31.7% (96/303) of the included studies assessed the effects on adherence to medication or other therapies, and 10% (30/303) assessed health literacy. The Danish MAST does not mention the measurement of health-related behavior changes [2]. Within the NICE framework, originally developed in the United Kingdom, applications with the purpose of improving health-related behaviors are assigned to their own group [3]. However, neither the MAST nor the NICE framework explicitly recommends capturing adherence or health literacy for all types of applications. Health literacy is not only an outcome but it is also a critical precondition for the successful use of telemedicine by the patient in addition to digital literacy. To ensure the appropriate use of the technology and the assessment of PROMs and PREMs, proper training and guidance of the users is of at least equal relevance, according to the literature [56,61-64]. Therefore, health literacy should not only be included in the evaluation merely for reasons of measuring effectiveness; it is also a possible factor influencing purposeful and successful telemedicine use by the patients [58,60,65,66]. In summary, future developments will show to what extent and in which way innovations from Germany will affect the use of PROMs and PREMs in evaluation studies of telemedicine applications.

Strengths and Limitations

One limitation of the study is that the medical purpose was classified by the ICD-10 chapters, all of which only describe a group of diseases and not the disease itself [20]. Chapter 1, for example, covers circulatory diseases, which include congenital heart defects, strokes, and aneurysms, all of which differ in etiology, symptoms, and therapy. There was a similar degree of heterogeneity in telemedicine applications. A more detailed distinction between user groups, setting, technical execution, and other criteria exists in the taxonomy used as a basis for the subgroup analysis, but this was not considered in our review [21]. The same applies to the analysis of single OMIs. The problem of heterogeneity is not an issue inherent only to this study. In their paper published in *Nature* in 2020, Guo et al [1] pointed out that the different types of interventions, medical purposes, and outcomes can lead to limitations in reviews of digital health interventions in general.

Another limitation was the large number of possible combinations of medical purpose, application, and study type. Nevertheless, several patterns were identified to answer the research questions, and the results of the entire sample could be transferred to use cases; thus, the influence of heterogeneity was not as great as initially assumed.

Another limitation might be that only 1 reviewer performed full-text screening. In the context of classical systematic reviews for the purpose of evidence synthesis of effectiveness or risk factors, screening by 2 reviewers is mandatory to minimize beta error. The approach of our review, on the other hand, was different. We intended to use the methodology of a systematic literature search to generate data for quantitative analysis. Owing to the 627 studies to be screened, an increased beta error in the form of missing studies seemed acceptable to us for reasons of research economics. As we wanted to conduct a plain descriptive analysis of the data with a total of 303 included studies, we did not consider the validity of the result to be compromised.

The strength of the review is that, to the best of our knowledge, this is the first systematic review investigating the characteristics of PROM and PREM use in evaluation studies of telemedicine applications covering all application types and medical purposes.

Reviews do exist for specific use cases; however, these usually do not cover all outcomes. Instead, they focus on selected outcomes for the purpose of evidence synthesis or do not focus exclusively on PROMs and PREMs [14-16,48,67-71].

Preliminary excerpts of the review results were presented to an expert audience of health care scientists at a conference in October 2020 [72].

Implications for Future Research

High heterogeneity reflected by the multitude of OMIs used per outcome domain and a lack of standardization poses a challenge to the selection of PROMs [70,71] and PREMs. New developments and updated versions of existing guidelines for the evaluation of telemedicine could contribute to further standardization in the selection of outcome domains and OMIs [73].

The use case analysis indicated that the most common outcome domains were HRQoL and emotional function, which could be the first starting point for further efforts. Equally, user satisfaction and usability [48] as well as health literacy and adherence [7] should be taken into account, although these outcome domains were not frequently surveyed in our review.

Further investigation will be required to reveal how the use of PROMs and PREMs for the evaluation of telemedicine will evolve over the next few years and if the trends observed in this review will persist.

In addition, upcoming studies will have to investigate how a greater consideration of PROMs and PREMs in German approval and reimbursement procedures for digital health applications will affect the future use of PROMs and PREMs in evaluation studies of telemedicine applications.

Conclusions

In recent years, there has been an increasing number of studies, particularly high-evidence studies, that use PROMs and PREMs to evaluate telemedicine services. Despite the great heterogeneity of telemedicine interventions and the associated evaluation approaches, several conclusions can be drawn. PROMs have been in the focus of evaluation studies. With the increasing maturity stage of telemedicine applications and higher evidence levels, the use of PROMs has increased. PREMs played a role, especially in the initial phases of application development, with low-evidence study designs. In this case, PREMs were primarily used to test the usability and acceptance of the application. Regardless of the findings, telemedicine-specific PREMs should be used more frequently and in a standardized manner to continuously evaluate

telemedicine service quality, both during and after implementation.

The distribution of the outcome domains showed that only HRQoL and emotional function were assessed in almost all studies. Simultaneously, health literacy as a precondition for using the application adequately, alongside proper training and guidance, has rarely been reported. At the level of the OMIs, it was shown that many different OMIs were used for each domain. Further efforts should be pursued for the standardization of PROM and PREM collection in evaluation studies of telemedicine applications.

Acknowledgments

The review was part of the project *Häusliche Gesundheitsstation*, which focused on the development of a telemonitoring system for patients with cardiac insufficiency in their home environment. In our work package, the current status of patient-centered evaluation of telemedicine needed to be investigated through a systematic review. The project was funded by the European Regional Development Fund and the Free State of Saxony (grant 100278533). The authors would like to thank the European Regional Development Fund and the Free State of Saxony for funding the project *Häusliche Gesundheitsstation* and making this review possible. We would also like to thank the partners of the project for their collaboration and joint activities within the project.

Authors' Contributions

AK, LH, JS, and MS were responsible for the study design of the review. AK coordinated the review. AK and LH conducted the searches. AK and SH extracted the data. AK conducted the analyses. AK and MS drafted the manuscript. AK, MS, LH, NE, SH, and JS critically evaluated the article and gave their final approval before submission.

Conflicts of Interest

JS received institutional grants for investigator-initiated trials from Novartis, Sanofi, ALK, and Pfizer. He acted as a consultant for Novartis, ALK, Lilly, and Sanofi.

Multimedia Appendix 1
Search string.
[DOCX File , 15 KB - jmir_v23i11e30042_app1.docx]

Multimedia Appendix 2
Study list.
[XLSX File (Microsoft Excel File), 93 KB - jmir_v23i11e30042_app2.xlsx]

References

1. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-A call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 34446821]
2. 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 [FREE Full text] [doi: 10.1017/S0266462311000638] [Medline: 22617736]
3. National Institute for Health and Care Excellence (NICE). Evidence standards framework for digital health technologies.: National Institute for Health and Care Excellence (NICE); 2021 Apr 23. URL: https://www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-pdf-1124017457605 [accessed 2021-07-27]
4. Arnold K, Scheibe M, Müller O, Schmitt J, und die CCS THOS Konsensgruppe. [Principles for the evaluation of telemedicine applications: Results of a systematic review and consensus process]. Z Evid Fortbild Qual Gesundhwes 2016 Nov;117:9-19 [FREE Full text] [doi: 10.1016/j.zefq.2016.04.011] [Medline: 27938734]
5. Kowatsch T, Otto L, Harpernink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. it - Inf Technol 2019 Nov 20;61(5-6):253-263 [FREE Full text] [doi: 10.1515/itit-2019-0019]
6. Khoja S, Durrani H, Scott RE, Sajwani A, Piryani U. Conceptual framework for development of comprehensive e-health evaluation tool. Telemed J E Health 2013 Jan;19(1):48-53 [FREE Full text] [doi: 10.1089/tmj.2012.0073] [Medline: 22957502]
7. Federal Institute for Drugs and Medical Devices (BfArM). The fast-track process for digital health applications (DiGA) according to section 139e SGBV. Bonn: Federal Institute for Drugs and Medical Devices (BfArM); 2020 Aug 07. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf;jsessionid=2C5B67183372C3B0DAAD7DC886953C1C.1_cid354?__blob=publicationFile&v=2 [accessed 2021-07-27]

8. Kidholm K, Clemensen J, Caffery LJ, Smith AC. The model for assessment of telemedicine (MAST): a scoping review of empirical studies. J Telemed Telecare 2017 Oct;23(9):803-813 [FREE Full text] [doi: 10.1177/1357633X17721815] [Medline: 28758525]
9. Nwe K, Larsen ME, Nelissen N, Wong DC. Medical mobile app classification using the national institute for health and care excellence evidence standards framework for digital health technologies: interrater reliability study. J Med Internet Res 2020 Jun 05;22(6):e17457 [FREE Full text] [doi: 10.2196/17457] [Medline: 32501271]
10. Locke HN, Brooks J, Arendsen LJ, Jacob NK, Casson A, Jones AK, et al. Acceptability and usability of smartphone-based brainwave entrainment technology used by individuals with chronic pain in a home setting. Br J Pain 2020 Aug;14(3):161-170 [FREE Full text] [doi: 10.1177/2049463720908798] [Medline: 32922777]
11. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools). Silver Spring: Food and Drug Administration; 2020 Nov 16. URL: https://www.fdanews.com/ext/resources/files/2020/11-24-20-BEST.pdf?1606261388 [accessed 2021-09-27]
12. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions, Second edition. Chichester: John Wiley & Sons; 2019:481.
13. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv Res 2019 Oct;54(5):1023-1035 [FREE Full text] [doi: 10.1111/1475-6773.13187] [Medline: 31218671]
14. Jackson BD, Gray K, Knowles SR, De Cruz P. EHealth technologies in inflammatory bowel disease: a systematic review. J Crohns Colitis 2016 Sep;10(9):1103-1121 [FREE Full text] [doi: 10.1093/ecco-jcc/jjw059] [Medline: 26928960]
15. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The use and effects of electronic health tools for patient self-monitoring and reporting of outcomes following medication use: systematic review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]
16. Bright P, Hambly K. What is the proportion of studies reporting patient and practitioner satisfaction with software support tools used in the management of knee pain and is this related to sample size, effect size, and journal impact factor? Telemed J E Health 2018 Aug 01;24(8):562-576 [FREE Full text] [doi: 10.1089/tmj.2017.0207] [Medline: 29265954]
17. Mackintosh A, Comabella C, Hadi M, Gibbons E, Roberts N, Fitzpatrick R. PROM group construct and instrument type filters. Department of Public Health, University of Oxford, Oxford. 2010 Feb. URL: https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf [accessed 2021-04-29]
18. Daliya P, Gemmill EH, Lobo DN, Parsons SL. A systematic review of patient reported outcome measures (PROMs) and quality of life reporting in patients undergoing laparoscopic cholecystectomy. Hepatobiliary Surg Nutr 2019 Jun;8(3):228-245 [FREE Full text] [doi: 10.21037/hbsn.2019.03.16] [Medline: 31245403]
19. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Gheorghe A, Keeley T, et al. Measurement properties of patient-reported outcome measures (PROMs) used in adult patients with chronic kidney disease: a systematic review. PLoS One 2017 Jun 21;12(6):e0179733 [FREE Full text] [doi: 10.1371/journal.pone.0179733] [Medline: 28636678]
20. World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision. Geneva: World Health Organization; 2011:1-195.
21. Harst L, Timpel P, Otto L, Richter P, Wollschlaeger B, Lantsch H. An empirically derived taxonomy of telemedicine - development of a standardized codebook. In: Proceedings of the 18th Deutscher Kongress für Versorgungsforschung. Berlin: German Medical Science Publishing House; 2019 Oct 02 Presented at: 18th Deutscher Kongress für Versorgungsforschung; October 09-11, 2019; Berlin, Germany URL: https://www.egms.de/static/en/meetings/dkvf2019/19dkvf024.shtml [doi: 10.3205/19dkvf024]
22. Harst L, Otto L, Timpel P, Richter P, Lantzsch H, Wollschlaeger B, et al. An empirically sound telemedicine taxonomy – applying the CAFE methodology. J Public Health (Berl.) 2021 May 28:s10389 [FREE Full text] [doi: 10.1007/s10389-021-01558-2]
23. Fitch C. Information systems in healthcare: mind the gap. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences.: IEEE; 2004 Feb 26 Presented at: 37th Annual Hawaii International Conference on System Sciences; Jan. 5-8, 2004; Big Island, HI, USA p. 5-8 URL: https://ieeexplore.ieee.org/document/1265367 [doi: 10.1109/hicss.2004.1265367]
24. Dierks C. Rechtliche und praktische probleme der integration von telemedizin — ein Problemaufriss. In: Rechtsfragen der Telemedizin. Berlin, Heidelberg: Springer; 2001:1-35.
25. Fong B, Fong A, Li C. Information technologies in medicine and digital health. In: Telemedicine Technologies: Information Technologies in Medicine and Digital Health. United States: John Wiley & Sons; May 29, 2020:1-297.
26. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011 Jul 18;17(6):484-494 [FREE Full text] [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
27. Schulz EG, Stahmann A, Neumann CL. Telemedicine: interventional decentralised blood pressure telemonitoring (idTBPM). Swiss Med Wkly 2015 Jan 11;145:w14077 [FREE Full text] [doi: 10.4414/smw.2015.14077] [Medline: 25587892]
28. Rogante M, Grigioni M, Cordella D, Giacomozzi C. Ten years of telerehabilitation: a literature overview of technologies and clinical applications. NeuroRehabilitation 2010 Dec 03;27(4):287-304 [FREE Full text] [doi: 10.3233/NRE-2010-0612] [Medline: 21160118]

29. Fitzner K, Moss G. Telehealth--an effective delivery method for diabetes self-management education? Popul Health Manag 2013 Jun 04;16(3):169-177 [FREE Full text] [doi: 10.1089/pop.2012.0054] [Medline: 23216062]
30. Sheridan N, Kenealy T, Kuluski K, McKillop A, Parsons J, Wong-Cornall C. Are patient and carer experiences mirrored in the practice reviews of self-management support (Prisms) provider taxonomy? Int J Integr Care 2017 Aug 27;17(2):8 [FREE Full text] [doi: 10.5334/ijic.2483] [Medline: 28970749]
31. Piga M, Cangemi I, Mathieu A, Cauli A. Telemedicine for patients with rheumatic diseases: systematic review and proposal for research agenda. Semin Arthritis Rheum 2017 Aug 01;47(1):121-128 [FREE Full text] [doi: 10.1016/j.semarthrit.2017.03.014] [Medline: 28420491]
32. Timpel P, Oswald S, Schwarz PE, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]
33. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 2018 Apr;96:84-92 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.12.020] [Medline: 29288712]
34. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Br Med J 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
35. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208 [FREE Full text] [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
36. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
37. 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]
38. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207 [FREE Full text] [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
39. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: a comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24-31 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
40. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). Med Care 2005 May;43(5):436-444 [FREE Full text] [doi: 10.1097/01.mlr.0000160375.47920.8c] [Medline: 15838407]
41. Gorst SA, Coates L. Telehealth acceptance measure: to assess patient motivation in the use of telehealth. The University of Manchester, Manchester. 2014 Nov 05. URL: http://malt.group.shef.ac.uk/assets/files/MALT%20TAM%20Final.pdf [accessed 2021-07-27]
42. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
43. 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 [FREE Full text] [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
44. Hirani SP, Rixon L, Beynon M, Cartwright M, Cleanthous S, Selva A, WSD Investigators. Quantifying beliefs regarding telehealth: development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. J Telemed Telecare 2017 May 25;23(4):460-469 [FREE Full text] [doi: 10.1177/1357633X16649531] [Medline: 27224997]
45. Demiris G, Speedie S, Finkelstein S. A questionnaire for the assessment of patients' impressions of the risks and benefits of home telecare. J Telemed Telecare 2000 Oct 01;6(5):278-284 [FREE Full text] [doi: 10.1258/1357633001935914] [Medline: 11070589]
46. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016 Jul 01;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
47. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B. Oxford centre for evidence-based medicine: levels of evidence (March 2009). Centre for Evidence-Based Medicine, University of Oxford. 2009 Mar. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 [accessed 2021-07-27]
48. Hajesmaeel-Gohari S, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. BMC Med Inform Decis Mak 2021 Feb 02;21(1):36 [FREE Full text] [doi: 10.1186/s12911-021-01407-y] [Medline: 33531013]
49. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007 May;45(5 Suppl 1):12-21 [FREE Full text] [doi: 10.1097/01.mlr.0000254567.79743.e2] [Medline: 17443114]

50. Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, et al. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. Health Serv Deliv Res 2016;4(16):55-68. [doi: 10.3310/hsdr04160] [Medline: 27227187]
51. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas 2018 Jul 11;9:353-367 [FREE Full text] [doi: 10.2147/PROM.S156279] [Medline: 30464666]
52. Kargo AS, Coulter A, Jensen PT, Steffensen KD. Proactive use of PROMs in ovarian cancer survivors: a systematic review. J Ovarian Res 2019 Jul 15;12(1):1-8 [FREE Full text] [doi: 10.1186/s13048-019-0538-9] [Medline: 31307510]
53. Siljander MP, McQuivey KS, Fahs AM, Galasso LA, Serdahely KJ, Karadsheh MS. Current trends in patient-reported outcome measures in total joint arthroplasty: a study of 4 major orthopaedic journals. J Arthroplasty 2018 Nov;33(11):3416-3421 [FREE Full text] [doi: 10.1016/j.arth.2018.06.034] [Medline: 30057269]
54. Weszl M, Rencz F, Brodszky V. Is the trend of increasing use of patient-reported outcome measures in medical device studies the sign of shift towards value-based purchasing in Europe? Eur J Health Econ 2019 Jun;20(Suppl 1):133-140 [FREE Full text] [doi: 10.1007/s10198-019-01070-1] [Medline: 31098886]
55. Steinbeck V, Ernst SC, Pross C. Patient-Reported Outcome Measures (PROMs): ein internationaler Vergleich. Bertellsmann Stiftung, Gütersloh. 2021 May 10. URL: https://www.bertelsmann-stiftung.de/en/publications/publication/did/patient-reported-outcome-measures-proms-ein-internationaler-vergleich [accessed 2021-09-27]
56. Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W. Capturing Patient-Reported Outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. Patient 2015 Aug;8(4):301-309 [FREE Full text] [doi: 10.1007/s40271-014-0090-z] [Medline: 25300613]
57. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3:1-6 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
58. Conard S. Best practices in digital health literacy. Int J Cardiol 2019 Oct 01;292:277-279 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.070] [Medline: 31230937]
59. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
60. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082 [FREE Full text] [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
61. Ly JJ, Crescioni M, Eremenco S, Bodart S, Donoso M, Butler AJ, et al. Training on the use of technology to collect patient-reported outcome data electronically in clinical trials: best practice recommendations from the ePRO consortium. Ther Innov Regul Sci 2019 Jul;53(4):431-440 [FREE Full text] [doi: 10.1177/2168479018796206] [Medline: 30221997]
62. Fleming S, Barsdorf AI, Howry C, O'Gorman H, Coons SJ. Optimizing electronic capture of clinical outcome assessment data in clinical trials: the case of patient-reported endpoints. Ther Innov Regul Sci 2015 Nov;49(6):797-804 [FREE Full text] [doi: 10.1177/2168479015609102] [Medline: 30222384]
63. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
64. Scheibe M, Lang C, Druschke D, Arnold K, Luntz E, Schmitt J, et al. Independent use of a home-based telemonitoring app by older patients with multimorbidity and mild cognitive impairment: qualitative study. JMIR Hum Factors 2021 Jul 12;8(3):e27156 [FREE Full text] [doi: 10.2196/27156] [Medline: 34255664]
65. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
66. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019 Oct;24(12):1668-1675 [FREE Full text] [doi: 10.1177/1359105317695429] [Medline: 28810415]
67. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient self-management of asthma using mobile health applications: a systematic review of the functionalities and effects. Appl Clin Inform 2017 Oct;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
68. Sul A, Lyu D, Park D. Effectiveness of telemonitoring versus usual care for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2020 May;26(4):189-199 [FREE Full text] [doi: 10.1177/1357633X18811757] [Medline: 30541375]
69. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28 [FREE Full text] [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
70. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]

71. Doshi H, Hsia B, Shahani J, Mowrey W, Jariwala SP. Impact of technology-based interventions on patient-reported outcomes in asthma: a systematic review. J Allergy Clin Immunol Pract 2021 Jun;9(6):2336-2341 [FREE Full text] [doi: 10.1016/j.jaip.2021.01.027] [Medline: 33548519]
72. Knapp A. Anwendung von PROMs und PREMs bei der evaluation von telemedizinischen anwendungen: überblick über die aktuelle praxis. In: Proceedings of the 19th Deutscher Kongress für Versorgungsforschung. Düsseldorf: German Medical Science Publishing House; 2020 Sep 25 Presented at: 19th Deutscher Kongress für Versorgungsforschung; Sept 30 - Oct 01, 2020; Berlin, Germany URL: https://www.egms.de/static/en/meetings/dkvf2020/20dkvf134.shtml [doi: 10.3205/20dkvf134]
73. Schlieter H, Timpel P, Otto L, Richter P, Wollschlaeger B, Knapp A, et al. Digitale Gesundheitsanwendungen – Forderungen für deren entwicklung, implementierung und begleitende evaluation. Monitor Versorgungsforschung 2021 Mar 30;14(02/2021):76-80 [FREE Full text] [doi: 10.24945/MVF.02.21.1866-0533.2301]

Abbreviations

AMSTAR 2: A Measurement Tool to Assess systematic Reviews
HRQoL: health-related quality of life
ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th revision
MAST: Model for Assessment of Telemedicine
NICE: National Institute for Health and Care Excellence
OMI: outcome measurement instrument
PREM: patient-reported experience measure
PROM: patient-reported outcome measure
RCT: randomized controlled trial

Edited by R Kukafka; submitted 29.04.21; peer-reviewed by S Hajesmaeel Gohari, H Mehdizadeh, S Newman, Y Jiang, S Guo; comments to author 15.06.21; revised version received 06.08.21; accepted 12.09.21; published 17.11.21.

Please cite as:
Knapp A, Harst L, Hager S, Schmitt J, Scheibe M
Use of Patient-Reported Outcome Measures and Patient-Reported Experience Measures Within Evaluation Studies of Telemedicine Applications: Systematic Review
J Med Internet Res 2021;23(11):e30042
URL: https://www.jmir.org/2021/11/e30042
doi:10.2196/30042
PMID:34523604

Review

Artificial Intelligence in Rehabilitation Targeting the Participation of Children and Youth With Disabilities: Scoping Review

Vera C Kaelin[1,2], MSc; Mina Valizadeh[3,4], BSc; Zurisadai Salgado[2,5], BSc; Natalie Parde[3,4*], PhD; Mary A Khetani[1,2,5,6*], ScD

[1]Rehabilitation Sciences, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

[2]Children's Participation in Environment Research Lab, University of Illinois at Chicago, Chicago, IL, United States

[3]Computer Science, College of Engineering, University of Illinois at Chicago, Chicago, IL, United States

[4]Natural Language Processing Laboratory, University of Illinois at Chicago, Chicago, IL, United States

[5]Occupational Therapy, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

[6]CanChild Centre for Childhood Disability Research, McMaster University, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Mary A Khetani, ScD
Occupational Therapy
College of Applied Health Sciences
University of Illinois at Chicago
1919 West Taylor Street, Room 316A
Chicago, IL, 60612-7250
United States
Phone: 1 312 996 0942
Email: mkhetani@uic.edu

Abstract

Background: In the last decade, there has been a rapid increase in research on the use of artificial intelligence (AI) to improve child and youth participation in daily life activities, which is a key rehabilitation outcome. However, existing reviews place variable focus on participation, are narrow in scope, and are restricted to select diagnoses, hindering interpretability regarding the existing scope of AI applications that target the participation of children and youth in a pediatric rehabilitation setting.

Objective: The aim of this scoping review is to examine how AI is integrated into pediatric rehabilitation interventions targeting the participation of children and youth with disabilities or other diagnosed health conditions in valued activities.

Methods: We conducted a comprehensive literature search using established Applied Health Sciences and Computer Science databases. Two independent researchers screened and selected the studies based on a systematic procedure. Inclusion criteria were as follows: participation was an explicit study aim or outcome or the targeted focus of the AI application; AI was applied as part of the provided and tested intervention; children or youth with a disability or other diagnosed health conditions were the focus of either the study or AI application or both; and the study was published in English. Data were mapped according to the types of AI, the mode of delivery, the type of personalization, and whether the intervention addressed individual goal-setting.

Results: The literature search identified 3029 documents, of which 94 met the inclusion criteria. Most of the included studies used multiple applications of AI with the highest prevalence of robotics (72/94, 77%) and human-machine interaction (51/94, 54%). Regarding mode of delivery, most of the included studies described an intervention delivered in-person (84/94, 89%), and only 11% (10/94) were delivered remotely. Most interventions were tailored to groups of individuals (93/94, 99%). Only 1% (1/94) of interventions was tailored to patients' individually reported participation needs, and only one intervention (1/94, 1%) described individual goal-setting as part of their therapy process or intervention planning.

Conclusions: There is an increasing amount of research on interventions using AI to target the participation of children and youth with disabilities or other diagnosed health conditions, supporting the potential of using AI in pediatric rehabilitation. On the basis of our results, 3 major gaps for further research and development were identified: a lack of remotely delivered participation-focused interventions using AI; a lack of individual goal-setting integrated in interventions; and a lack of interventions tailored to individually reported participation needs of children, youth, or families.

(J Med Internet Res 2021;23(11):e25745) doi:10.2196/25745

KEYWORDS

health care; pediatric rehabilitation; technology; young persons; robotics; human-machine interaction; personalization; customization; goal-setting; natural language processing; machine learning

Introduction

Background

Technology-based interventions are of increased importance in pediatric rehabilitation and can be useful to rehabilitation practitioners when delivering family-centered and function-focused interventions to service-eligible children, youth, and families [1]. In addition, technology-based rehabilitation tools can be useful to organizations that have electronic data capture systems to monitor trends in rehabilitation service use and outcomes for quality improvement [2,3]. For both individuals and organizations, the COVID-19 pandemic has heightened the demand for technological solutions to remotely deliver and monitor rehabilitation services [4].

One way to provide technology-based pediatric rehabilitation is by applying artificial intelligence (AI), which is a priority of the National Institutes of Health, as reflected in their Rehabilitation Research Plan [5,6]. According to Russell and Norvig [7], AI is concerned with designing and building systems that think like humans, act like humans, think rationally, and act rationally. It encompasses different subfields such as natural language processing (NLP), robotics, or human augmentics [7]. The application of AI in pediatric rehabilitation has the potential to simplify steps in the therapeutic process and possibly decrease provider and patient burden as well as afford providers to customize their rehabilitation services.

Rehabilitation includes a broad range of highly variable interventions that are challenging to define owing to their complexity [8-10]. One important way to classify rehabilitation intervention is through its targeted outcome [9]. In the last decade, there has been a rapid increase in research on the use of AI to improve key pediatric rehabilitation outcomes, including body functions, activity performance, and the full participation of children and youth with disabilities in valued activities [11-13]. For children and youth, participation in home, school, and community activities has been defined by the World Health Organization as “involvement in life situations” [14] and was further conceptualized by Imms et al [15] as attendance and involvement in activities, which is related to but distinct from their activity competencies, environment or context, and their preferences or sense of self [15]. Given the unmet participation need among children or youth with disabilities and other diagnosed health conditions, beginning in early childhood and across settings [16-19], there is a growing number of participation-focused intervention studies [20], including interventions that integrate AI to target the participation of young persons receiving pediatric rehabilitation.

A recent systematic literature review on the effect of participation-focused pediatric rehabilitation identified 2257 records through a database search, indicating the high relevance of participation as an outcome in pediatric rehabilitation interventions [20]. However, this review does not focus on AI use. Literature reviews focusing on AI indicate that the use of AI in the form of information and communication technology or robots may improve children’s engagement in play, stimulate school performance [13], and promote social interactions [11,12]. However, these reviews place variable focus on participation [11-13], are narrow in scope (eg, focus on participation in play only) [13], and are restricted to select diagnoses (eg, physical disability) [13]. These limitations hinder our understanding of the existing scope of AI applications in pediatric rehabilitation that target participation in daily life activities.

Objectives

To better understand the current scope of AI applications within pediatric rehabilitation and to identify gaps for future research, there is a critical need to summarize existing evidence on the use of AI across interventions targeting child and youth participation in activities. The purpose of this scoping review is to examine how AI is integrated into pediatric rehabilitation interventions targeting the participation of children and youth with disabilities or other diagnosed health conditions in valued activities.

Our paper’s contributions are as follows:

1. An overview of the scope of literature focusing on AI targeting participation as a primary pediatric rehabilitation outcome and top priority from the perspective of families.
2. A summary of the types of AI and personalization used in the interventions over a time span of more than 20 years.
3. Identification of research gaps based on the found and summarized literature with a focus on AI targeting the participation of children and youth with disabilities or other diagnosed health conditions.

Methods

Design

Scoping reviews are commonly used to provide an overview of existing evidence in a certain field and identify gaps for future research [21,22]. The increasing number of publications on the use of AI in participation-focused pediatric rehabilitation indicates an emerging field for the advancement of rehabilitation research and therefore justifies the need to conduct this scoping review [21]. The protocol for this scoping review was registered in the Open Science Framework [23].

Search Strategy

The first author of this review (VCK) conducted a systematic literature search in well-established databases in the fields of Applied Health Sciences and Computer Science (PubMed, PsycINFO, ERIC, CINAHL, IEEE Xplore, and ACM Digital Library) for documents published before February 2021. No

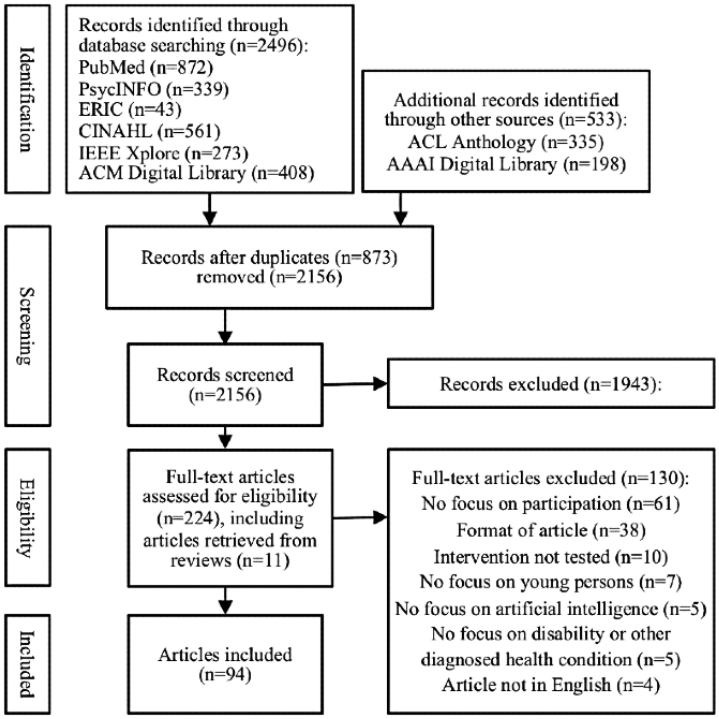
other publication data limit or search limitations were applied to the search. We solicited support from a health sciences librarian to develop subject headings for each database with available thesaurus (ie, PubMed, PsycINFO, ERIC, and CINAHL) and keywords for *artificial intelligence*, *participation*, *health care*, *disability*, and *young persons* (Textbox 1) [24,25]. These were applied using truncations and Boolean terms, resulting in 2496 documents (Figure 1). The full search strategy is presented in Multimedia Appendix 1. After consultation with an AI expert (NP), additional searches were performed in ACL Anthology and AAAI Digital Library, using the same keywords from the database searches. This led to an additional 533 documents.

Textbox 1. Search strategy.

Main search term and additional search terms for abstract and title search

- artificial intelligence
  - affective computing, algorithms, chatbot, cognitive computing, computer vision, constraint optimization, constraint satisfaction, data mining, data processing, deep learning, expert systems, feature extraction, fuzzy logic, game theory, human computation, image analysis, inductive logic programming, knowbot*, knowledge bases, knowledge-based agent, knowledge engineering, knowledge representation, machine learning, natural language processing, neural networks, pattern recognition, predictive model, reinforcement learning, robot*, semantic networks, semi-supervised learning, supervised learning, text analysis, unsupervised learning, virtual agent, virtual reality
- participation
  - attendance, engag*, inclus*, involvement
- health care
  - health care, healthcare, rehabilitation, therap*
- disability
  - disab*, handicap*, impair*, special needs, special need
- young persons
  - adolesc*, caregiv*, child*, family, families, infant*, paediatric*, parent*, pediatric*, student*, teen*, toddler*, young adult, young adults, youth*

Figure 1. Study selection process.



Screening and Selection Process

After removal of duplicates, 2 independent coders (VCK and MV) applied inclusion and exclusion criteria to the title and abstracts of the remaining documents. Documents indicating potential fit based on their abstract or title underwent full-text reading and were coded based on the same inclusion and exclusion criteria. First, 2 independent authors with expertise in participation-focused pediatric rehabilitation (VCK and ZS) coded the same documents until at least 80% agreement was reached [26]. Discrepancies and coding uncertainties were resolved through discussion. The remaining documents were screened separately (VCK and ZS), whereas 20% of randomly selected documents underwent double screening (ie, 10%: VCK and ZS; 10% including an external reviewer with expertise in participation-focused pediatric rehabilitation: VCK and Kyle A Truevillian) [26]. Doubts regarding document inclusion were discussed with a third reviewer (VCK, ZS, and Kyle A Truevillian). Second, documents indicating fit were further screened by an additional author (MV) with a specific focus on AI. Disagreements were resolved through discussion (VCK, ZS, MV, and Kyle A Truevillian) and with the help of additional key informants (MAK and NP).

Documents were included if (1) participation was an explicit study aim or outcome, or the targeted focus of the AI application; (2) AI was applied as part of the provided and tested intervention; (3) children or youth with a disability or other diagnosed health condition were the focus of either the study, AI application or both; and (4) the study was published in English. To ensure the inclusion of a broad scope of studies, no operational definition of participation was used when applying the selection criteria. Studies or AI interventions focused on participation; inclusion; engagement; playfulness; access to, or attendance in life situations, settings or activities; social interaction; or social engagement were considered participation and were therefore included in this review. Documents were excluded if (1) participation in daily activities was not the focus of the study (eg, focus was on skill development); (2) there was no use of AI for the described intervention (eg, the term *algorithm* was used in a noncomputer science way); (3) interventions using AI were not tested with either children, youth or both; (4) there was no focus on disability or other diagnosed health condition; (5) studies focused on adults (mean age >24 years [27]); (6) the study was published in languages other than English; or (7) documents were textbooks or textbook reviews, literature reviews, study protocols, conference or workshop programs, or only abstracts without additional information. To prevent missing relevant studies, each reference list of the excluded literature reviews was screened.

Data Extraction and Analysis

The data extraction template using Microsoft Excel was trialed with 5 studies by the first author and discussed with the author team to ensure the clarity and relevance of the included categories for data extraction [28]. Data were then extracted for all included studies based on the following categories: authors, year, title, sample size, child and youth age, child and youth race and ethnicity, family socioeconomic status, parental education, the types of AI applied in the intervention, the intervention's type of personalization, whether the primary method for intervention delivery was in-person or remote, and whether goal-setting was addressed as part of the intervention.

The mapping of included studies to one or multiple types of AI was guided by CSRankings [29] and the AAAI keywords taxonomy [30], 2 commonly used ranking systems and taxonomies for AI. The types of AI in this study include robotics, NLP, human computation and crowdsourcing, computer vision, knowledge representation and reasoning, machine learning (ML), human-machine interaction (HMI), cognitive modeling, constraint satisfaction and optimization, game theory, planning and routing and scheduling, and visualization and virtual reality (VR) [7]. As robotic devices are increasingly used in rehabilitation [31], studies that used robotics were further classified according to whether they focused on robot mechanics or on the system of use.

A framework developed by Fan et al [32] guided the mapping of the intervention's type of personalization according to 2 dimensions (ie, categorical vs individuated personalization; implicit vs explicit personalization). Categorical personalization targets a category of individuals, such as a diagnostic group or single-child families [32]. For this review, these were, for example, devices that are designed to include features that meet the common needs of children with autism spectrum disorder (ASD). Individuated personalization targets specific individuals [32], in the case of this review, individually perceived and reported participation needs. Implicit personalization is system-initiated, meaning it is automatically done by the system, whereas explicit personalization is user-initiated, meaning users manually guide the system on the preferred adaptation [32].

Interventions were mapped as in-person when the intervention was delivered face-to-face with a researcher or rehabilitation professional. Interventions were considered as remotely delivered when they were conducted in the child's natural environment and without a researcher or rehabilitation professional.

Mapping of included studies with regard to whether goal-setting was addressed as part of the described rehabilitation services was guided by the goal-setting and action-planning practice framework for rehabilitation settings [33]. Studies were mapped to address goal-setting if the described rehabilitation services included goal negotiation (ie, where the patient is at and where the patient would like to get to) or goal-setting (ie, what the patient would like to achieve) [33].

Charted data were summarized using descriptive statistics (ie, frequency counts and percentages) to provide an overview of the available evidence on how AI is used to support participation among children and youth with disabilities or other diagnosed health conditions.

Results

Overview of Found and Included Research

The literature search revealed 3029 documents with 873 duplicates (ie, documents appeared multiple times), resulting in 2156 documents entering the 2-fold screening process to

assess their eligibility based on the inclusion and exclusion criteria (Figure 1). The first screening phase included titles and abstracts and led to 213 included and 1943 excluded documents, as well as 11 additional studies found when screening the reference list of excluded literature reviews. The Cohen κ for interrater agreement was 0.67, indicating a substantial agreement [34]. This estimate did not include the numerous conference programs (n=450) found through AAAI Digital Library and ACL Anthology, for which determining exclusion was trivial, resulting in a more conservative Cohen κ value.

The second screening phase included a full-text review of the 224 included documents from the first screening phase, resulting in 94 included studies for this scoping review. Of the 130 excluded documents, 61 (46.9%) lacked focus on participation, 38 (29.2%) were excluded because of their format (ie, textbook or textbook review, study protocol, literature review, or only abstract), 10 (7.7%) did not test the intervention, and 7 (5.4%) addressed an adult population; 5 (3.8%) did not use AI in the intervention, 5 (3.8%) did not focus on people with disability or other diagnosed health conditions, and 4 (3.1%) were not written in English (Figure 1).

Type of Included Research

The 94 included studies were published between 2000 and 2021, with a higher proportion published after 2010 (76/94, 81%; Multimedia Appendix 2 [35-128]). All studies included AI as part of their intervention and targeted children or youth participation, as described in their research aims, outcomes, or as their focus of the tested AI application.

As for sample characteristics, the described interventions were evaluated on sample sizes ranging from 1 to 120 children and youth with an average of 14 children or youth. Of the included studies that reported on gender identity, 76% (51/67) had a higher proportion of boys represented in their sample. A total of 92% (86/94) of the included studies did not report on the socioeconomic background of the family, parental education, or child or youth race or ethnicity. In total, 15% (14/94) of studies sampled caregivers, teachers, peers, other school staff, or combinations thereof, in addition to children or youth when evaluating the intervention. Included interventions were developed or tested for children or youth with a broad range of diagnoses, with ASD being the most prevalent (43/94, 46%), followed by cerebral palsy (CP; 18/94, 19%).

Types of AI Intervention, Mode of Intervention Delivery, and Type of Personalization

Most of the 94 included studies used robotics as the type of AI intervention to target participation among children and youth with disabilities or other diagnosed health conditions (72/94, 77%) [35-106], followed by HMI (51/94, 54%) [35, 37-44, 47, 50-55, 58, 65, 66, 69, 71, 72, 75, 78, 80, 82, 85-89, 91-98, 100, 101, 103-111], visualization and VR (19/94, 20%) [53,54,72,79,107,108,112-124], NLP (18/94, 19%) [36,47,52,64,71,78,79,82,91,101,103-105,120,125-128], ML (11/94, 12%), computer vision (10/94, 11%) [40,41,64,67,69,107,110,120,125,126,128], and constraint satisfaction and optimization (1/94, 1%; Table 1) [110]. Of the 72 studies on robotics, 63 (88%) studies focused on the system of use [35-39, 41-53, 55, 57, 58, 60-63, 65, 66, 68-74, 76-78, 80-101, 103-106] and 9 (13%) focused on robot mechanics [40,54,56,59,64,67,75,79,102].

Table 1. Delivery of participation-focused rehabilitation interventions that include artificial intelligence (AI).

Type of AI	Personalization	Mode of delivery	Addresses individual goal-setting
Robotics: 72 [35-106]	• Implicit + individuated: 0 • Implicit + categorical: 37 [35, 36, 38, 43-46, 48-50, 52, 55-57, 60, 64-66, 70, 74-77, 79, 81-84, 87, 89-91, 97, 101, 104-106] • Explicit + individuated: 0 • Explicit + categorical: 35 [37, 39-42, 47, 51, 53, 54, 58, 59, 61-63, 67-69, 71-73, 78, 80, 85, 86, 88, 92-96, 98-100, 102, 103]	• In-person: 67 [35-50,52-84,86-91,93,94,96-99,101-106] • Remote: 5 [51,85,92,95,100]	0
Human-machine interaction: 51 [35, 37-44, 47, 50-55, 57, 58, 65, 66, 69, 71, 72, 75, 78, 80, 82, 85-89, 91-98, 100, 101, 103-111]	• Implicit + individuated: 1 [110] • Implicit + categorical: 21 [35, 38, 43, 44, 50, 52, 55, 57, 65, 66, 75, 82, 87, 89, 91, 97, 101, 104-107] • Explicit + individuated: 0 • Explicit + categorical: 29 [37, 39-42, 47, 51, 53, 54, 58, 69, 71, 72, 78, 80, 85, 86, 88, 92-96, 98, 100, 103, 108, 109, 111]	• In-person: 44 [35, 37-44, 47, 50, 52-55, 57, 58, 65, 66, 69, 71, 72, 75, 78, 80, 82, 86-89, 91, 93, 94, 96-98, 101, 103-107, 109, 111] • Remote: 7 [51,85,92,95,100,108,110]	0
Visualization and virtual reality: 19 [53,54,72,79,107,108,112-124]	• Implicit + individuated: 0 • Implicit + categorical: 5 [79,107,114,117,120] • Explicit + individuated: 0 • Explicit + categorical: 14 [53,54,72,108,112,113,115,116,118,119,121-124]	• In-person: 18 [53,54,72,79,107,112-124] • Remote: 1 [108]	1 [112]
Natural language processing: 18 [36, 47, 52, 64, 71, 78, 79, 82, 91, 101, 103-105, 120, 125-128]	• Implicit + individuated: 0 • Implicit + categorical: 14 [36,52,64,79,82,91,101,104,105,120,125-128] • Explicit + individuated: 0 • Explicit + categorical: 4 [47,71,78,103]	• In-person: 15 [36,47,52,64,71,78,79,82,91,101,103-105,120,127] • Remote: 3 [125,126,128]	0
Machine learning: 11 [40,41,64,67,69,107,110,120,125,126,128]	• Implicit + individuated: 1 [110] • Implicit + categorical: 6 [64,107,120,125,126,128] • Explicit + individuated: 0 • Explicit + categorical: 4 [40,41,67,69]	• In-person: 7 [40,41,64,67,69,107,120] • Remote: 4 [110,125,126,128]	0
Computer vision: 10 [35,39,58,63,65,69,75,112,120,127]	• Implicit + individuated: 0 • Implicit + categorical: 5 [35,65,75,120,127] • Explicit + individuated: 0 • Explicit + categorical: 5 [39,58,63,69,112]	• In-person: 10 [35,39,58,63,65,69,75,112,120,127] • Remote: 0	1 [112]
Constraint satisfaction and optimization: 1 [110]	• Implicit + individuated: 1 [110] • Implicit + categorical: 0 • Explicit + individuated: 0 • Explicit + categorical: 0	• In-person: 0 • Remote: 1 [110]	0
Human computation and crowdsourcing: 0	N/A[a]	N/A	N/A
Planning, routing, and scheduling: 0	N/A	N/A	N/A
Cognitive modeling: 0	N/A	N/A	N/A
Game theory: 0	N/A	N/A	N/A

[a]N/A: not applicable.

Most of the included studies described interventions using multiple applications of AI (60/94, 64%), such as robotics with HMI [35-44, 47, 50-55, 57, 58, 63-67, 69, 71, 72, 75, 78-80, 82, 85-89, 91-98, 100, 101, 103-108, 112, 120, 125-128], or

ML with NLP and constraint satisfaction and optimization [110]. Across these studies, robotics was most often integrated into interventions that employed multiple applications of AI. Examples of multiple AI interventions that include robotics are humanoid or nonhumanoid devices to facilitate interaction or play of children with disabilities by directing the robot head toward a target or rocking its body from left to right to express emotions such as excitement [39].

Out of the included 94 studies, 22 (23%) studies used forms of AI other than robotics [107-128]. Of these, 15 included visualization and VR applications, such as an immersive virtual learning program [107,108,112-124]; 8% (7/94) of interventions included neither robotics nor visualization and VR [109-111,125-128]. Examples of such interventions are a framework for speech-to-sign language translation for children with hearing impairments [127] and the design of a virtual space for hospitalized children to meet with their peers [108].

As for mode of delivery, most of the included studies described an intervention delivered in-person (84/94, 89%) [35-50,52-84,86-91,93,94,96-99,101-107,109,111-124,127], mainly using a one-on-one approach. A total of 11% (10/94) of included studies evaluated an AI intervention that was delivered remotely [51,85,92,95,100,108,110,125,126,128].

Most AI interventions were tailored to a category of individuals (ie, categorical personalization) such as by a diagnostic group (93/94, 99%) [35-109,111-128], using implicit (ie, automatically personalized: 45/94, 48%) [35, 36, 38, 43-46, 48-50, 52, 55-57, 60, 64-66, 70, 74-77, 79, 81-84, 87, 89-91, 97, 101, 104-107, 114, 117, 120, 125-128] or explicit (ie, manually personalized: 48/94, 51%) approaches [37, 39-42, 47, 51, 53, 54, 58, 59, 61-63, 67-69, 71-73, 78, 80, 85, 86, 88, 92-96, 98-100, 102, 103, 108, 109, 111-113, 115, 116, 118, 119, 121-124]. For example, Yee et al [35] designed a robotic platform for children with ASD by tailoring it to the needs typically described by this diagnostic group. In contrast, only 1% (1/94) of the included studies described an intervention that was tailored to the individually reported and unique needs of the child or youth with a disability or other diagnosed health condition (ie, individuated personalization) [110]. It included the use of a recommender algorithm, integrating information about the location of different physical and virtual learning resources, their purposes, modality, as well as the individual's class schedules, university rooms, and navigation system to suggest suitable and uniquely tailored options for access and navigation to the appropriate location [110]. In addition, 1% (1/94) of the included interventions described individual goal-setting as part of their therapy process or intervention planning [112]. In this intervention, the Canadian Occupational Performance Measure was used for individual goal-setting and a video game–based task-oriented activity training was performed according to the defined patient goal [112].

Discussion

Principal Findings

This study summarizes 2 decades of evidence on the use of AI across interventions targeting the participation of children and youth with disabilities or other diagnosed health conditions, extending knowledge on the breadth of using AI in pediatric rehabilitation. There is an increased interest in AI applications for customizing pediatric rehabilitation services to individual child and family reported needs and reducing provider burden. The results of this review suggest that AI applications designed for children of diverse ages and diagnoses tend to emphasize robotics (alone or in combination with other forms of AI), in-person delivery, and targeted groups of children using implicit and explicit personalization approaches. Each finding is further discussed to identify knowledge gaps that warrant future research.

Most of the studied robotic devices are not commercially available and were used during on-site therapy sessions to *train* a child or youth to participate in a specific activity, with an expected transfer or carryover of that gain into the child or youth's natural environment of home, school, or community. This expectation has been challenged in previous participation literature, emphasizing the importance of environments for shaping a young person's participation in daily activities [129-132]. The mediating role of the environment and context for child and youth participation has also been supported in research examining the effect of participation-focused interventions [133,134].

Interestingly, most of the found interventions were delivered in-person, despite the potential for leveraging technology to deliver rehabilitation interventions remotely. This result is in line with a previously conducted survey, which indicated that only 8% of Americans used telemedicine in 2019 [135]. Alternatively, our results might also be due to the high prevalence of interventions using robotics, often requiring the presence of trained operators and specialized equipment on the therapy site [48,52,84]. Remotely delivered interventions using robotics deploy robots in classrooms to enable virtual inclusion of home-bound children [51,85,92,95,100]. The remaining remotely delivered interventions commonly apply ML and NLP, potentially indicating the suitability of ML and NLP for use in remote pediatric rehabilitation interventions using AI. ML and NLP have been used in a range of health interventions to promote behavioral changes, such as physical activity and healthy diet, including goal-setting [136,137]. Given the existing evidence on the use of AI for goal-setting in other health care domains [136,137] and the importance of gaining efficiency in enacting the complex process of goal-setting in pediatric rehabilitation [138,139], the lack of attention to goal-setting in this review indicates a clear knowledge gap warranting future research. Emerging electronic participation–focused interventions such as the Participation and Environment Measure–Plus [140-143] with individual goal-setting as an integral part of their intervention might benefit from exploring the use of AI to fill this knowledge gap.

Most of the identified AI applications were tailored to the needs of groups of individuals, with only 1% (1/94) being tailored to the individually reported participation needs of children and youth with disabilities or other diagnosed health conditions. When comparing this result with the use of AI in fields outside of health care, it is surprising. For example, in marketing, AI has revolutionized common advertisement practices by tailoring

advertisements to the reported needs and preferences of clients. This discrepancy between fields might be due to stricter protection of health information; however, there is an increase in similar advancements using data collected from patients in formal (eg, electronic health records) or informal (eg, patient dialog) settings, such as for diagnosing and decision-making [144-148]. Similar approaches might also be possible and beneficial within pediatric rehabilitation, using existing patient data to predict tailored participation–focused interventions. A recent systematic literature review on the effects of participation-focused interventions recommends focusing on individually tailored interventions to support the participation of children and youth with disabilities [20]. One way to tailor rehabilitation interventions to the reported needs of patients involves the patient's goals. In rehabilitation, goal-setting has become an integral part of the therapy process across professions, including pediatric participation–focused interventions [133,134]. Previous research has shown that caregivers can be guided to create participation-focused goals on the web [149]. Including goal-setting in AI-supported pediatric rehabilitation interventions might be an important first step to enable tailoring interventions to the participation needs of children and youth with disabilities or other diagnosed health conditions.

Despite the high prevalence of included studies testing or designing interventions for children and youth with ASD or cerebral palsy, a diverse sample in terms of diagnoses was represented in this scoping review, indicating relevance for the use of AI applications across diagnoses. In contrast, only 9% (8/94) of studies reported on child or youth race or ethnicity, family socioeconomic status, parental education, or family income, despite evidence indicating its influence on child and youth participation [150-152]. Future research should capture child and youth race and ethnicity as well as indicators of socioeconomic family status to describe the diversity of their study sample [153].

Limitations

An effort was made to conduct a comprehensive review of the literature pertaining to the use of AI to target children and youth participation. However, the results of this scoping review should be interpreted in light of some limitations. Despite the relatively high number of included studies, we may have missed some relevant documents. Three primary examples include (1) if an intervention using AI was not identified as such during the screening of titles and abstracts, the document was likely excluded from the search or selection process; (2) screening of reference lists was undertaken for review articles versus all included studies; and (3) documents published in languages other than English were excluded. In addition, the included studies were not screened based on their definition of participation, potentially leading to conceptual inconsistency, as has been shown in a systematic review of participation-focused interventions for children with disabilities [20]. Variability in the conceptualization of participation can limit the interpretation and comparison of results across studies [20,154] to identify knowledge gaps specific to participation-focused rehabilitation interventions. Future research should map studies using AI to contemporary frameworks of the participation concept to ensure the interpretability of results across studies.

Conclusions

There is an increasing amount of research on interventions using AI to target the participation of children and youth with disabilities or other diagnosed health conditions, supporting the potential of using AI in pediatric rehabilitation. Overall, most interventions used multiple AI applications, including robotics and HMI. Other types of AI, such as ML or NLP, were less prevalent but showed potential benefits in participation-focused intervention. On the basis of our results, 3 major gaps were identified, warranting the need for future research and development: (1) a lack of remotely provided participation-focused interventions using AI; (2) a lack of individual goal-setting integrated in interventions using AI; and (3) a lack of interventions using AI tailored to individually reported participation needs of children, youth, or families.

In addition, future research should consistently report on the socioeconomic background of the family, parental education, or race and ethnicity to describe the diversity of their study sample.

Acknowledgments

The authors acknowledge Kyle A Truevillian for assisting with full-text screening of documents for this review and Amelia Brunskill at the University of Illinois at Chicago library for guidance on search strategies for undertaking this review. The authors also thank Vivian Villegas and Varun Maheshwari from the Children's Participation and Environment Research Lab at the University of Illinois at Chicago for critical feedback on prior drafts. This work was conducted in partial fulfillment of the requirements for a PhD in Rehabilitation Sciences and was funded by the University of Illinois at Chicago, through their Dean's Scholar Fellowship (principal investigator: VCK) and Chancellor's Undergraduate Research Award (ZS). We acknowledge the Research Open Access Publishing Fund of the University of Illinois at Chicago for financial support toward the open access publishing fee for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search history in the included databases.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i11e25745_app1.pdf]

Multimedia Appendix 2
Included studies.
[PDF File (Adobe PDF File), 98 KB - jmir_v23i11e25745_app2.pdf]

References

1. Wang S, Blazer D, Hoenig H. Can ehealth technology enhance the patient-provider relationship in rehabilitation? Arch Phys Med Rehabil 2016 Sep;97(9):1403-1406. [doi: 10.1016/j.apmr.2016.04.002] [Medline: 27109332]
2. Lowes LP, Noritz GH, Newmeyer A, Embi PJ, Yin H, Smoyer WE, 'Learn from Every Patient' Study Group. 'Learn From Every Patient': implementation and early results of a learning health system. Dev Med Child Neurol 2017 Feb 22;59(2):183-191 [FREE Full text] [doi: 10.1111/dmcn.13227] [Medline: 27545839]
3. Stevenson RD. Integration of research and clinical practice: the future is now. Dev Med Child Neurol 2017 Feb 16;59(2):119-120 [FREE Full text] [doi: 10.1111/dmcn.13281] [Medline: 27634036]
4. Cacioppo M, Bouvier S, Bailly R, Houx L, Lempereur M, Mensah-Gourmel J, ECHO Group. Emerging health challenges for children with physical disabilities and their parents during the COVID-19 pandemic: The ECHO French survey. Ann Phys Rehabil Med 2021 Aug 18;64(3):101429 [FREE Full text] [doi: 10.1016/j.rehab.2020.08.001] [Medline: 32818674]
5. Eunice Kennedy Shriver National Institute of Child Health and Human Development and the NIH Medical Rehabilitation Coordinating Committee. NIH Rehabilitation Plan 2020. URL: https://nichd.ideascalegov.com/a/campaign-home/51 [accessed 2021-01-21]
6. Eunice Kennedy Shriver National Institute of Child Health and Human Development and the NIH Medical Rehabilitation Coordinating Committee. NIH Rehabilitation Plan. Eunice Kennedy Shriver National Institute of Child Health and Human Development and the NIH Medical Rehabilitation Coordinating Committee. 2020. URL: https://nichd.ideascalegov.com/a/campaign-home/51 [accessed 2021-01-21]
7. Russell S, Norvig P. Artificial Intelligence: A Modern Approach, 3rd Ed. New York: Pearson; 2015:1-1152.
8. Jette AM. Opening the black box of rehabilitation interventions. Phys Ther 2020 Jun 23;100(6):883-884. [doi: 10.1093/ptj/pzaa078] [Medline: 32585702]
9. Whyte J, Hart T. It's more than a black box; it's a Russian doll. Am J Phys Med Rehabil 2003 Aug;82(8):639-652. [doi: 10.1097/01.phm.0000078200.61840.2d] [Medline: 12872021]
10. Dejong G, Horn SD, Gassaway JA, Slavin MD, Dijkers MP. Toward a taxonomy of rehabilitation interventions: using an inductive approach to examine the "black box" of rehabilitation. Arch Phys Med Rehabil 2004 Apr;85(4):678-686. [doi: 10.1016/j.apmr.2003.06.033] [Medline: 15083447]
11. Sartorato F, Przybylowski L, Sarko DK. Improving therapeutic outcomes in autism spectrum disorders: enhancing social communication and sensory processing through the use of interactive robots. J Psychiatr Res 2017 Jul;90:1-11. [doi: 10.1016/j.jpsychires.2017.02.004] [Medline: 28213292]
12. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012 Aug 15;14(1):275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]
13. van den Heuvel RJF, Lexis MAS, Gelderblom GJ, Jansens RML, de Witte LP. Robots and ICT to support play in children with severe physical disabilities: a systematic review. Disabil Rehabil Assist Technol 2016 Feb 02;11(2):103-116. [doi: 10.3109/17483107.2015.1079268] [Medline: 26330097]
14. World Health Organization. International Classification of Functioning, Disability and Health: Children and Youth Version: ICF-CY. Geneva: World Health Organization; 2007:1-322.
15. Imms C, Granlund M, Wilson PH, Steenbergen B, Rosenbaum PL, Gordon AM. Participation, both a means and an end: a conceptual analysis of processes and outcomes in childhood disability. Dev Med Child Neurol 2017 Jan 19;59(1):16-25 [FREE Full text] [doi: 10.1111/dmcn.13237] [Medline: 27640996]
16. Kaelin VC, Wallace ER, Werler MM, Collett BR, Khetani MA. Community participation in youth with craniofacial microsomia. Disabil Rehabil 2020 Jun 01:1-8. [doi: 10.1080/09638288.2020.1765031] [Medline: 32478589]
17. Kaelin VC, Wallace ER, Werler MM, Collett BR, Rosenberg J, Khetani MA. Caregiver perspectives on school participation among students with craniofacial microsomia. Am J Occup Ther 2021 Jan 28;75(2):1-10 [FREE Full text] [doi: 10.5014/ajot.2021.041277] [Medline: 33657352]
18. Khetani MA, Graham JE, Davies PL, Law MC, Simeonsson RJ. Psychometric properties of the young children's participation and environment measure. Arch Phys Med Rehabil 2015 Feb;96(2):307-316 [FREE Full text] [doi: 10.1016/j.apmr.2014.09.031] [Medline: 25449189]
19. Law M, Anaby D, Teplicky R, Khetani MA, Coster W, Bedell G. Participation in the home environment among children and youth with and without disabilities. Br J Occup Ther 2013 Feb 15;76(2):58-66. [doi: 10.4276/030802213x13603244419112]

20. Adair B, Ullenhag A, Keen D, Granlund M, Imms C. The effect of interventions aimed at improving participation outcomes for children with disabilities: a systematic review. Dev Med Child Neurol 2015 Dec 24;57(12):1093-1104 [FREE Full text] [doi: 10.1111/dmcn.12809] [Medline: 26010935]
21. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Australia: JBI - The University of Adelaide; 2020.
22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/m18-0850]
23. Kaelin VC, Valizadeh M, Salgado Z, Truevillian KA, Parde M, Khetani MA. Artificial intelligence in rehabilitation targeting participation of young persons with disabilities: a scoping review. Open Science Forum. 2020. URL: https://osf.io/wmkg8 [accessed 2021-09-21]
24. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). In: Cochrane Training. Hoboken, New Jersey, United States: Wiley–Blackwell; 2021. URL: www.training.cochrane.org/handbook
25. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York: CRD, University of York; 2009.
26. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 22;19:160940691989922. [doi: 10.1177/1609406919899220]
27. United Nations. Youth. URL: https://www.un.org/en/global-issues/youth [accessed 2020-05-24]
28. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
29. Berger E. CSRankings: Computer science rankings. 2020. URL: http://csrankings.org/#/index?all [accessed 2021-09-18]
30. Conitzer V. Association for the Advancement of Artificial Intelligence. AAAI-19: Thirty-third AAAI conference on artificial intelligence 2019 Jul 17;33:9755-9759 [FREE Full text] [doi: 10.1609/aaai.v33i01.33019755]
31. Riener R. Rehabilitation robotics. FNT in Robotics 2013;3(1-2):1-137. [doi: 10.1561/2300000028]
32. Fan H, Poole MS. What is personalization? Perspectives on the design and implementation of personalization in information systems. J Organ Comput Electron Commer 2006 Jan;16(3-4):179-202. [doi: 10.1080/10919392.2006.9681199]
33. Scobbie L, Dixon D, Wyke S. Goal setting and action planning in the rehabilitation setting: development of a theoretically informed practice framework. Clin Rehabil 2011 May 03;25(5):468-482. [doi: 10.1177/0269215510389198] [Medline: 21131335]
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
35. Yee AWH, Kee TY, Limbu DK, Jian ATH, Dung TA, Yuen AWC. Developing a robotic platform to play with pre-school autistic children in a classroom environment. In: Proceedings of the Workshop at SIGGRAPH Asia. 2012 Presented at: SA '12: SIGGRAPH Asia 2012; Nov 26 - 27, 2012; Singapore p. 81-86. [doi: 10.1145/2425296.2425311]
36. So W, Cheng C, Lam W, Huang Y, Ng K, Tung H, et al. A robot-based play-drama intervention may improve the joint attention and functional play behaviors of chinese-speaking preschoolers with autism spectrum disorder: a pilot study. J Autism Dev Disord 2020 Feb 26;50(2):467-481. [doi: 10.1007/s10803-019-04270-z] [Medline: 31655965]
37. Klein T, Gelderblom GJ, de Witte L, Vanstipelen S. Evaluation of short term effects of the IROMEC robotic toy for children with developmental disabilities. In: Proceedings of the IEEE International Conference on Rehabilitation Robotics. 2011 Presented at: IEEE International Conference on Rehabilitation Robotics; Jun 29 - July 1, 2011; Zurich, Switzerland p. 1-5. [doi: 10.1109/icorr.2011.5975406]
38. Ríos-Rincón AM, Adams K, Magill-Evans J, Cook A. Playfulness in children with limited motor abilities when using a robot. Phys Occup Ther Pediatr 2016 Nov 13;36(3):232-246. [doi: 10.3109/01942638.2015.1076559] [Medline: 26566226]
39. Kozima H, Nakagawa C, Yasuda Y. Children–robot interaction: a pilot study in autism therapy. Prog Brain Res 2007;164:385-400. [doi: 10.1016/S0079-6123(07)64021-7]
40. Javed H, Park CH. Interactions with an empathetic agent: regulating emotions and improving engagement in autism. IEEE Robot Automat Mag 2019 Jun;26(2):40-48. [doi: 10.1109/mra.2019.2904638]
41. Javed H, Burns R, Jeon M, Howard AM, Park CH. A robotic framework to facilitate sensory experiences for children with autism spectrum disorder: a preliminary study. ACM Trans Hum Robot Interact 2019 Dec 31;9(1):1-26 [FREE Full text] [doi: 10.1145/3359613] [Medline: 33829148]
42. Iacono I, Lehmann H, Marti P, Robins B, Dautenhahn K. Robots as social mediators for children with autism - a preliminary analysis comparing two different robotic platforms. In: Proceedings of the IEEE International Conference on Development and Learning (ICDL). 2011 Presented at: IEEE International Conference on Development and Learning (ICDL); Aug 24-27, 2011; Frankfurt am Main, Germany. [doi: 10.1109/devlrn.2011.6037322]
43. Goodrich MA, Colton MA, Brinton B, Fujiki M. A case for low-dose robotics in autism therapy. In: Proceedings of the 6th International Conference on Human-Robot Interaction. 2011 Presented at: HRI'11: International Conference on Human-Robot Interaction; Mar 6 - 9, 2011; Lausanne, Switzerland p. 143-144. [doi: 10.1145/1957656.1957702]

44. Gilliaux M, Renders A, Dispa D, Holvoet D, Sapin J, Dehez B, et al. Upper limb robot-assisted therapy in cerebral palsy: a single-blind randomized controlled trial. Neurorehabil Neural Repair 2015 Feb 11;29(2):183-192. [doi: 10.1177/1545968314541172] [Medline: 25015650]
45. Cook AM, Bentz B, Harbottle N, Lynch C, Miller B. School-based use of a robotic arm system by children with disabilities. IEEE Trans Neural Syst Rehabil Eng 2005 Dec;13(4):452-460. [doi: 10.1109/tnsre.2005.856075]
46. Conchinha C, Osório P, de Freitas JC. Playful learning: educational robotics applied to students with learning disabilities. In: Proceedings of the 2015 International Symposium on Computers in Education (SIIE). 2015 Presented at: International Symposium on Computers in Education (SIIE); Nov 25-27, 2015; Setubal, Portugal p. 167-171. [doi: 10.1109/siie.2015.7451669]
47. Silvera-Tawil D, Bradford D, Roberts-Yates C. Talk to me: the role of human-robot interaction in improving verbal communication skills in students with autism or intellectual disability. In: Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2018 Presented at: 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); Aug 27-31, 2018; Nanjing, China. [doi: 10.1109/roman.2018.8525698]
48. Bonarini A, Clasadonte F, Garzotto F, Gelsomini M, Romero M. Playful interaction with Teo, a mobile robot for children with neurodevelopmental disorders. In: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2016 Presented at: DSAI 2016: 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion; Dec 1 - 3, 2016; Vila Real Portugal p. 223-231. [doi: 10.1145/3019943.3019976]
49. Besio S, Bonarini A, Lynch H, Molina P, Veronese F, Bulgarelli D. Mainstream robotic toys and children with physical impairment: what about playfulness? In: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2016 Presented at: DSAI 2016: 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion; Dec 1 - 3, 2016; Vila Real Portugal p. 208-215. [doi: 10.1145/3019943.3019974]
50. Aslam S, Standen PJ, Shopland N, Burton A, Brown D. A comparison of humanoid and non-humanoid robots in supporting the learning of pupils with severe intellectual disabilities. In: Proceedings of the International Conference on Interactive Technologies and Games (ITAG). 2016 Presented at: International Conference on Interactive Technologies and Games (ITAG); Oct 26-27, 2016; Nottingham, UK p. 7-12. [doi: 10.1109/itag.2016.9]
51. Newhart VA, Warschauer M, Sender L. Virtual inclusion via telepresence robots in the classroom: an exploratory case study. Int J Technol Learn 2016;23(4):9-25. [doi: 10.18848/2327-0144/cgp/v23i04/9-25]
52. Adams K, Cook A. Using robots in "Hands-on" academic activities: a case study examining speech-generating device use and required skills. Disabil Rehabil Assist Technol 2016 Dec 11;11(5):433-443. [doi: 10.3109/17483107.2014.986224] [Medline: 25495803]
53. Welch KC, Lahiri U, Warren Z, Sarkar N. An approach to the design of socially acceptable robots for children with autism spectrum disorders. Int J of Soc Robotics 2010 Jul 7;2(4):391-403. [doi: 10.1007/s12369-010-0063-x]
54. Wainer J, Dautenhahn K, Robins B, Amirabdollahian F. A pilot study with a novel setup for collaborative play of the humanoid robot Kaspar with children with autism. Int J of Soc Robotics 2013 Sep 11;6(1):45-65. [doi: 10.1007/s12369-013-0195-x]
55. Stanton CM, Kahn PH, Severson RL, Ruckert JH, Gill BT. Robotic animals might aid in the social development of children with autism. In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction. 2008 Presented at: HRI '08: International Conference on Human Robot Interaction; Mar 12 - 15, 2008; Amsterdam, The Netherlands p. 271-278. [doi: 10.1145/1349822.1349858]
56. Pulido JC, González JC, Suárez-Mejías C, Bandera A, Bustos P, Fernández F. Evaluating the child–robot interaction of the naotherapist platform in pediatric rehabilitation. Int J of Soc Robotics 2017 Apr 8;9(3):343-358. [doi: 10.1007/s12369-017-0402-2]
57. Pliasa S, Fachantidis N. Can a robot be an efficient mediator in promoting dyadic activities among children with Autism Spectrum Disorders and children of Typical Development? In: Proceedings of the 9th Balkan Conference on Informatics. 2019 Presented at: BCI'19: 9th Balkan Conference in Informatics; Sept 26 - 28, 2019; Sofia, Bulgaria p. 1-6. [doi: 10.1145/3351556.3351592]
58. Rudovic O, Utsumi Y, Lee J, Hernandez J, Castelló Ferrer E, Schuller B, et al. Culturenet: a deep learning approach for engagement intensity estimation from face images of children with autism. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2018 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); Oct 1-5, 2018; Madrid, Spain p. 339-346. [doi: 10.1109/iros.2018.8594177]
59. Marti P. Bringing playfulness to disabilities. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. 2010 Presented at: NordiCHI '10: 6th Nordic Conference on Human-Computer Interaction; Oct 16 - 20, 2010; Reykjavik, Iceland p. 851-856. [doi: 10.1145/1868914.1869046]
60. Ludi S, Bernstein D, Mutch-Jones K. Enhanced robotics!: improving building and programming learning experiences for students with visual impairments. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education.

2018 Presented at: SIGCSE '18: The 49th ACM Technical Symposium on Computer Science Education; Feb 21 - 24, 2018; Baltimore, USA p. 372-377. [doi: 10.1145/3159450.3159501]
61. Lindsay S, Lam A. Exploring types of play in an adapted robotics program for children with disabilities. Disabil Rehabil Assist Technol 2018 Apr 28;13(3):263-270. [doi: 10.1080/17483107.2017.1306595] [Medline: 28350526]
62. Lindsay S, Hounsell KG. Adapting a robotics program to enhance participation and interest in STEM among children with disabilities: a pilot study. Disabil Rehabil Assist Technol 2017 Oct 20;12(7):694-704. [doi: 10.1080/17483107.2016.1229047] [Medline: 27762644]
63. Kozima H, Michalowski MP, Nakagawa C. Keepon: a playful robot for research, therapy, and entertainment. Int J of Soc Robotics 2009 Nov 19;1(1):3-18. [doi: 10.1007/s12369-008-0009-8]
64. Kokkoni E, Mavroudi E, Zehfroosh A, Galloway JC, Vidal R, Heinz J, et al. GEARing smart environments for pediatric motor rehabilitation. J Neuroeng Rehabil 2020 Feb 10;17(1):1-15 [FREE Full text] [doi: 10.1186/s12984-020-0647-0] [Medline: 32041623]
65. Koch SA, Stevens CE, Clesi CD, Lebersfeld JB, Sellers AG, McNew ME, et al. A feasibility study evaluating the emotionally expressive robot SAM. Int J of Soc Robotics 2017 Jun 23;9(4):601-613. [doi: 10.1007/s12369-017-0419-6]
66. Kim ES, Daniell CM, Makar C, Elia J, Scassellati B, Shic F. Potential clinical impact of positive affect in robot interactions for autism intervention. In: Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII). 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction (ACII); Sept 21-24, 2015; Xi'an, China. [doi: 10.1109/acii.2015.7344544]
67. Khamassi M, Chalvatzaki G, Tsitsimis T, Velentzas G, Tzafestas C. A framework for robot learning during child-robot interaction with human engagement as reward signal. In: Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2018 Presented at: 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); Aug 27-31, 2018; Nanjing, China p. 461-464. [doi: 10.1109/roman.2018.8525598]
68. Jung SE, Lee K, Cherniak S, Cho E. Non-sequential learning in a robotics class: insights from the engagement of a child with autism spectrum disorder. Tech Know Learn 2019 Jan 23;25(1):63-81. [doi: 10.1007/s10758-018-9394-8]
69. Pioggia G, Igliozzi R, Sica ML, Ferro M, Muratori F, Ahluwalia A, et al. Exploring emotional and imitational android-based interactions in autistic spectrum disorders. J CyberTher Rehabil 2008;1(1):49-61 [FREE Full text]
70. Haumont T, Rahman T, Sample W, M King M, Church C, Henley J, et al. Wilmington robotic exoskeleton: a novel device to maintain arm improvement in muscular disease. J Pediatr Orthop 2011 ;31(5):e44-e49. [doi: 10.1097/BPO.0b013e31821f50b5] [Medline: 21654447]
71. Ferm UM, Claesson BK, Ottesjö C, Ericsson S. Participation and enjoyment in play with a robot between children with cerebral palsy who use AAC and their peers. Augment Altern Commun 2015 Jun 29;31(2):108-123. [doi: 10.3109/07434618.2015.1029141] [Medline: 25921358]
72. Encarnação P, Leite T, Nunes C, Nunes da Ponte M, Adams K, Cook A, et al. Using assistive robots to promote inclusive education. Disabil Rehabil Assist Technol 2017 May 26;12(4):352-372. [doi: 10.3109/17483107.2016.1167970] [Medline: 27115833]
73. Cook A, Howery K, Gu J, Meng M. Robot enhanced interaction and learning for children with profound physical disabilities. Technol Disabil 2001 May 14;13(1):1-8. [doi: 10.3233/tad-2000-13101]
74. Buitrago JA, Bolaños AM, Caicedo Bravo E. A motor learning therapeutic intervention for a child with cerebral palsy through a social assistive robot. Disabil Rehabil Assist Technol 2020 Apr 26;15(3):357-362. [doi: 10.1080/17483107.2019.1578999] [Medline: 30806105]
75. Boccanfuso L, O'Kane JM. CHARLIE : an adaptive robot design with hand and face tracking for use in autism therapy. Int J of Soc Robotics 2011 Sep 24;3(4):337-347. [doi: 10.1007/s12369-011-0110-2]
76. Blanson Henkemans OA, Bierman BPB, Janssen J, Looije R, Neerincx MA, van Dooren MMM, et al. Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. Int J Hum Comput Stud 2017 Oct;106:63-76. [doi: 10.1016/j.ijhcs.2017.06.001]
77. Blanson Henkemans OA, Bierman BPB, Janssen J, Neerincx MA, Looije R, van der Bosch H, et al. Using a robot to personalise health education for children with diabetes type 1: a pilot study. Patient Educ Couns 2013 Aug;92(2):174-181. [doi: 10.1016/j.pec.2013.04.012] [Medline: 23684366]
78. Barakova EI, Bajracharya P, Willemsen M, Lourens T, Huskens B. Long-term LEGO therapy with humanoid robot for children with ASD. Expert Syst 2014 Nov 13;32(6):698-709. [doi: 10.1111/exsy.12098]
79. Arpaia P, Duraccio L, Moccaldi N, Rossi S. Wearable brain–computer interface instrumentation for robot-based rehabilitation by augmented reality. IEEE Trans Instrum Meas 2020 Sep;69(9):6362-6371. [doi: 10.1109/tim.2020.2970846]
80. Marti P, Giusti L. A robot companion for inclusive games: a user-centred design perspective. In: Proceedings of the IEEE International Conference on Robotics and Automation. 2010 Presented at: IEEE International Conference on Robotics and Automation; May 3-7, 2010; Anchorage, US p. 4348-4353. [doi: 10.1109/robot.2010.5509385]
81. Alotaibi MH. The effectiveness of robot-assisted, task-specific ankle training in improving deficits across the three domains of the ICF in children with cerebral palsy (CP). Thesis and Dissertations - Indiana University. 2018. URL: https://scholarworks.iupui.edu/handle/1805/17929 [accessed 2021-09-18]

82. Santatiwongchai S, Kaewkamnerdpong B, Jutharee W, Ounjai W. BLISS: using robot in learning intervention to promote social skills for autism therapy. In: Proceedings of the International Convention on Rehabilitation Engineering & Assistive Technology. 2016 Presented at: International Convention on Rehabilitation Engineering & Assistive Technology; Jul 25 - 28, 2016; Midview City, Singapore p. 1-4. [doi: 10.5555/3014393.3014413]
83. Schroeder AS, Homburg M, Warken B, Auffermann H, Koerte I, Berweck S, et al. Prospective controlled cohort study to evaluate changes of function, activity and participation in patients with bilateral spastic cerebral palsy after robot-enhanced repetitive treadmill therapy. Eur J Paediatr Neurol 2014 Jul;18(4):502-510. [doi: 10.1016/j.ejpn.2014.04.012] [Medline: 24821475]
84. Bulgarelli D, Bianquin N, Besio S, Molina P. Children with cerebral palsy playing with mainstream robotic toys: playfulness and environmental supportiveness. Front Psychol 2018 Sep 26;9:1-9 [FREE Full text] [doi: 10.3389/fpsyg.2018.01814] [Medline: 30319509]
85. Soares N, Kay JC, Craven G. Mobile robotic telepresence solutions for the education of hospitalized children. Perspect Health Inf Manag 2017;14(Fall):1e [FREE Full text] [Medline: 29118682]
86. Alnajjar F, Cappuccio M, Renawi A, Mubin O, Loo CK. Personalized robot interventions for autistic children: an automated methodology for attention assessment. Int J of Soc Robotics 2020 Mar 06;13(1):67-82. [doi: 10.1007/s12369-020-00639-8]
87. Barnes JA, Park CH, Howard A, Jeon M. Child-robot interaction in a musical dance game: an exploratory comparison study between typically developing children and children with autism. Int J Hum Comput Interact 2021 Sep 14;37(3):249-266. [doi: 10.1080/10447318.2020.1819667] [Medline: 33767571]
88. Metatla O, Bardot S, Cullen M, Serrano M, Jouffrais C. Robots for inclusive play- co-designing an educational game with visually impaired and sighted children. In: Proceedings 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, USA p. 1-13. [doi: 10.1145/3313831.3376270]
89. Ismail LI, Hanapiah FA, Belpaeme T, Dambre J, Wyffels F. Analysis of attention in child–robot interaction among children diagnosed with cognitive impairment. Int J of Soc Robotics 2020 Feb 03;13(2):141-152. [doi: 10.1007/s12369-020-00628-x]
90. Silva K, Lima M, Fafiães C, Sinval J, de Sousa L. Preliminary test of the potential of contact with dogs to elicit spontaneous imitation in children and adults with severe autism spectrum disorder. Am J Occup Ther 2020 Dec 05;74(1):1-8 [FREE Full text] [doi: 10.5014/ajot.2020.031849] [Medline: 32078518]
91. Marti P, Pollini A, Rullo A, Shibata T. Engaging with artificial pets. In: Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics. 2005 Presented at: Annual Conference on European Association of Cognitive Ergonomics; Sept 29 - October 1, 2005; Chania, Greece p. 99-106. [doi: 10.5555/1124666.1124680]
92. Weibel M, Nielsen MK, Topperzer MK, Hammer NM, Møller SW, Schmiegelow K, et al. Back to school with telepresence robot technology: a qualitative pilot study about how telepresence robots help school-aged children and adolescents with cancer to remain socially and academically connected with their school classes during treatment. Nurs Open 2020 Jul 12;7(4):988-997 [FREE Full text] [doi: 10.1002/nop2.471] [Medline: 32587717]
93. Neto I, Johal W, Couto M, Nicolau H, Paiva A, Guneysu A. Using tabletop robots to promote inclusive classroom experiences. In: Proceedings of the Interaction Design and Children Conference. 2020 Presented at: IDC '20: Interaction Design and Children; Jun 21 - 24, 2020; London, UK p. 281-292. [doi: 10.1145/3392063.3394439]
94. Costa S, Santos C, Soares F, Ferreira M, Moreira F. Promoting interaction amongst autistic adolescents using robots. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; Aug 31 - Sept. 4, 2010; Buenos Aires, Argentina p. 3856-3859. [doi: 10.1109/iembs.2010.5627905]
95. Fels DI, Waalen JK, Zhai S, Weiss PT. Telepresence under exceptional circumstances: enriching the connection to school for sick children. In: Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction. 2001 Presented at: IFIP TC13 International Conference on Human-Computer Interaction; Jul 9-13, 2001; Tokyo, Japan p. 617-624 URL: https://dblp.org/rec/conf/interact/FelsWZW01.html
96. Kim E, Paul R, Shic F, Scassellati B. Bridging the research gap: making HRI useful to individuals with autism. J Human-Robot Interact 2012 Aug 01;1(1):26-54. [doi: 10.5898/jhri.1.1.kim]
97. Shamsuddin S, Yussof H, Ismail LI, Mohamed S, Hanapiah FA, Zahari NI. Initial response in HRI - a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot NAO. Procedia Eng 2012;41:1448-1455. [doi: 10.1016/j.proeng.2012.07.334]
98. Simut RE, Vanderfaeillie J, Peca A, Van de Perre G, Vanderborght B. Children with autism spectrum disorders make a fruit salad with probo, the social robot: an interaction study. J Autism Dev Disord 2016 Jan 25;46(1):113-126. [doi: 10.1007/s10803-015-2556-9] [Medline: 26304031]
99. Feng H. Xylo-Bot: a therapeutic robot-based music platform for children with autism. ProQuest Dissertations - University of Denver. 2020. URL: https://www.proquest.com/openview/16a69b9a20d3fc4d0929951844b2faf8/1?pq-origsite=gscholar&cbl=44156 [accessed 2021-09-20]
100. Ahumada-Newhart V, Olson JS. Going to school on a robot: robot and user interface design features that matter. ACM Trans Comput Hum Interact 2019 Jul 27;26(4):1-28 [FREE Full text] [doi: 10.1145/3325210] [Medline: 31692962]

101. Miyamoto E, Lee M, Fujii H, Okada M. How can robots facilitate social interaction of children with autism?: possible implications for educational environments. In: Proceedings of the Fifth International Workshop on Epigenetic Robotics. 2005 Presented at: Fifth International Workshop on Epigenetic Robotics; 2005; Lund, Sweden p. 145-146 URL: http://cogprints.org/4993/1/miyamoto.pdf
102. Ludi S, Reichlmayr T. The use of robotics to promote computing to pre-college students with visual impairments. ACM Trans Comput Educ 2011 Oct;11(3):1-20. [doi: 10.1145/2037276.2037284]
103. Zhanatkyzy A, Telisheva Z, Turarova A, Zhexenova Z, Sandygulova A. Quantitative results of robot-assisted therapy for children with autism, ADHD and delayed speech development. In: Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. 2020 Presented at: HRI '20: ACM/IEEE International Conference on Human-Robot Interaction; Mar 23 - 26, 2020; Cambridge, UK p. 541-542. [doi: 10.1145/3371382.3378254]
104. Butchart J, Harrison R, Ritchie J, Martí F, McCarthy C, Knight S, et al. Child and parent perceptions of acceptability and therapeutic value of a socially assistive robot used during pediatric rehabilitation. Disabil Rehabil 2021 Jan 23;43(2):163-170. [doi: 10.1080/09638288.2019.1617357] [Medline: 31120794]
105. Ljunglöf P, Claesson B, Müller I, Ericsson S, Ottesjö C, Berman A, et al. Lekbot: a talking and playing robot for children with disabilities. In: Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies. 2011 Presented at: Second Workshop on Speech and Language Processing for Assistive Technologies; Jul 30, 2011; Edinburgh, Scotland p. 110-119. [doi: 10.5555/2140499.2140516]
106. Lehmann H, Iacono I, Robins B, Marti P, Dautenhahn K. Make it move': playing cause and effect games with a robot companion for children with cognitive disabilities. In: Proceedings of the 29th Annual European Conference on Cognitive Ergonomics. 2011 Presented at: ECCE '11: European Conference on Cognitive Ergonomics; Aug 24 - 26, 2011; Rostock, Germany p. 105-112. [doi: 10.1145/2074712.2074734]
107. Bian D, Wade J, Swanson A, Weitlauf A, Warren Z, Sarkar N. Design of a physiology-based adaptive virtual reality driving platform for individuals with ASD. ACM Trans Access Comput 2019 Feb 23;12(1):1-24. [doi: 10.1145/3301498]
108. Bers MU, Gonzalez-Heydrich J, DeMaso DR. Identity construction environments: supporting a virtual therapeutic community of pediatric patients undergoing dialysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001 Presented at: CHI01: Human Factors in Computing Systems; Mar 31 - Apr 5, 2001; Seattle, USA p. 380-387. [doi: 10.1145/365024.365302]
109. Porayska-Pomsta K, Alcorn AM, Avramides K, Beale S, Bernardini S, Foster ME, et al. Blending human and artificial intelligence to support autistic children's social communication skills. ACM Trans Comput-Hum Interact 2018 Dec 20;25(6):1-35. [doi: 10.1145/3271484]
110. Goldberg M, Karimi H, Pearlman JL. Interactive, Mobile, AGIle and Novel Education (IMAGINE): a conceptual framework to support students with mobility challenges in higher education. Disabil Rehabil Assist Technol 2016 Sep 12;11(1):50-60. [doi: 10.3109/17483107.2014.959074] [Medline: 25212492]
111. Kolne K, Bui S, Lindsay S. Assessing the environmental quality of an adapted, play-based LEGO robotics program to achieve optimal outcomes for children with disabilities. Disabil Rehabil 2020 Mar 25:1-10. [doi: 10.1080/09638288.2020.1743776] [Medline: 32208981]
112. Arman N, Tarakci E, Tarakci D, Kasapcopur O. Effects of video games–based task-oriented activity training (Xbox 360 Kinect) on activity performance and participation in patients with juvenile idiopathic arthritis. Am J Phys Med Rehabil 2019 Mar;98(3):174-181. [doi: 10.1097/phm.0000000000001001]
113. Reid DT. Benefits of a virtual play rehabilitation environment for children with cerebral palsy on perceptions of self-efficacy: a pilot study. Pediatr Rehabil 2002 Jul 10;5(3):141-148. [doi: 10.1080/1363849021000039344] [Medline: 12581476]
114. Miller S, Reid D. Doing play: competency, control, and expression. Cyberpsychol Behav 2003 Dec;6(6):623-632. [doi: 10.1089/109493103322725397] [Medline: 14756927]
115. Bonney E, Ferguson G, Smits-Engelsman B. The efficacy of two activity-based interventions in adolescents with developmental coordination disorder. Res Dev Disabil 2017 Dec;71:223-236. [doi: 10.1016/j.ridd.2017.10.013] [Medline: 29055242]
116. Green D, Wilson PH. Use of virtual reality in rehabilitation of movement in children with hemiplegia-a multiple case study evaluation. Disabil Rehabil 2012 Oct 06;34(7):593-604. [doi: 10.3109/09638288.2011.613520] [Medline: 21978233]
117. Mineo BA, Ziegler W, Gill S, Salkin D. Engagement with electronic screen media among students with autism spectrum disorders. J Autism Dev Disord 2009 Jan 15;39(1):172-187. [doi: 10.1007/s10803-008-0616-0] [Medline: 18626761]
118. Ke F, Im T. Virtual-reality-based social interaction training for children with high-functioning autism. J Educ Res 2013 Nov 02;106(6):441-461. [doi: 10.1080/00220671.2013.832999]
119. Reid D. The influence of virtual reality on playfulness in children with cerebral palsy: a pilot study. Occup Ther Int 2004 Aug;11(3):131-144. [doi: 10.1002/oti.202] [Medline: 15297894]
120. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism. IEEE Trans Neural Syst Rehabil Eng 2013 Jan;21(1):55-64. [doi: 10.1109/tnsre.2012.2218618]
121. Altizer Jr R, Handman E, Bayles G, Jackman J, Cheng K, Ritchie S, et al. Choreografish: co-designing a choreography-based therapeutic virtual reality system with youth who have autism spectrum advantages. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. 2018 Presented at: CHI PLAY '18:

The annual symposium on Computer-Human Interaction in Play; Oct 28 - 31, 2018; Melbourne, Australia p. 381-389. [doi: 10.1145/3270316.3271541]
122. Mei C, Mason L, Quarles J. 'I Built It!' - Exploring the effects of customizable virtual humans on adolescents with ASD. In: Proceedings of the IEEE Virtual Reality (VR). 2015 Presented at: IEEE Virtual Reality (VR); Mar 23-27, 2015; Arles, France p. 235-236. [doi: 10.1109/vr.2015.7223382]
123. Bortone I, Leonardis D, Solazzi M, Procopio C, Crecchi A, Bonfiglio L, et al. Integration of serious games and wearable haptic interfaces for neuro rehabilitation of children with movement disorders: a feasibility study. In: Proceedings of the International Conference on Rehabilitation Robotics (ICORR). 2017 Presented at: International Conference on Rehabilitation Robotics (ICORR); Jul 17-20, 2017; London, UK p. 1094-1099. [doi: 10.1109/icorr.2017.8009395]
124. Byrne J, Ip HHS, Lau K, Chen Li R, Tso A, Choi C. InSPAL: a novel immersive virtual learning programme. Stud Health Technol Inform 2015;219:129-134. [Medline: 26799893]
125. Rivas-Perez T, Meza-Vega A, Mora-Lezcano V, Palacio-Amador D, Chacón-Rivas M. EULER - Mathematical editing by voice input for people with visual impairment. In: Proceedings of the International Conference on Inclusive Technologies and Education (CONTIE). 2019 Presented at: International Conference on Inclusive Technologies and Education (CONTIE); Oct 30 - Nov. 1, 2019; San Jose del Cabo, Mexico p. 9-95. [doi: 10.1109/contie49246.2019.00012]
126. Black R, Waller A, Turner R, Reiter E. Supporting personal narrative for children with complex communication needs. ACM Trans Comput-Hum Interact 2012 Jul;19(2):1-35. [doi: 10.1145/2240156.2240163]
127. Otoom M, Alzubaidi MA. Ambient intelligence framework for real-time speech-to-sign translation. Assist Technol 2018 Feb 02;30(3):119-132. [doi: 10.1080/10400435.2016.1268218] [Medline: 28152342]
128. Beaudry J, Consigli A, Clark C, Robinson KJ. Getting ready for adult healthcare: designing a chatbot to coach adolescents with special health needs through the transitions of care. J Pediatr Nurs 2019 Nov;49:85-91. [doi: 10.1016/j.pedn.2019.09.004] [Medline: 31644960]
129. Anaby D, Hand C, Bradley L, DiRezze B, Forhan M, DiGiacomo A, et al. The effect of the environment on participation of children and youth with disabilities: a scoping review. Disabil Rehabil 2013 Sep 25;35(19):1589-1598. [doi: 10.3109/09638288.2012.748840] [Medline: 23350759]
130. Anaby D, Law M, Hanna S, Dematteo C. Predictors of change in participation rates following acquired brain injury: results of a longitudinal study. Dev Med Child Neurol 2012 Apr;54(4):339-346 [FREE Full text] [doi: 10.1111/j.1469-8749.2011.04204.x] [Medline: 22257122]
131. Albrecht EC, Khetani MA. Environmental impact on young children's participation in home-based activities. Dev Med Child Neurol 2017 Apr 17;59(4):388-394 [FREE Full text] [doi: 10.1111/dmcn.13360] [Medline: 27988938]
132. Houtrow A, Murphy N, Council on Children With Disabilities. Prescribing physical, occupational, and speech therapy services for children with disabilities. Pediatrics 2019 Apr 25;143(4):e20190285 [FREE Full text] [doi: 10.1542/peds.2019-0285] [Medline: 30910917]
133. Anaby DR, Law M, Feldman D, Majnemer A, Avery L. The effectiveness of the Pathways and Resources for Engagement and Participation (PREP) intervention: improving participation of adolescents with physical disabilities. Dev Med Child Neurol 2018 May 06;60(5):513-519 [FREE Full text] [doi: 10.1111/dmcn.13682] [Medline: 29405282]
134. Anaby DR, Law MC, Majnemer A, Feldman D. Opening doors to participation of youth with physical disabilities: an intervention study. Can J Occup Ther 2016 Apr 21;83(2):83-90. [doi: 10.1177/0008417415608653] [Medline: 26501799]
135. Telehealth index: 2019 Consumer survey. American Well Corporation. 2019. URL: https://static.americanwell.com/app/uploads/2019/07/American-Well-Telehealth-Index-2019-Consumer-Survey-eBook2.pdf [accessed 2021-03-16]
136. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
137. Gupta I, Eugenio B, Ziebart B, Liu B, Gerber B, Sharp L. Modeling health coaching dialogues for behavioral goal extraction. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2019 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 18-21, 2019; San Diego, USA p. 118-119. [doi: 10.1109/bibm47256.2019.8983138]
138. Pritchard-Wiart L, Thompson-Hodgetts S, McKillop AB. A review of goal setting theories relevant to goal setting in paediatric rehabilitation. Clin Rehabil 2019 Sep 09;33(9):1515-1526. [doi: 10.1177/0269215519846220] [Medline: 31072123]
139. Brewer K, Pollock N, Wright FV. Addressing the challenges of collaborative goal setting with children and their families. Phys Occup Ther Pediatr 2014 May 15;34(2):138-152. [doi: 10.3109/01942638.2013.794187] [Medline: 23672252]
140. Jarvis JM, Gurga A, Greif A, Lim H, Anaby D, Teplicky R, et al. Usability of the Participation and Environment Measure Plus (PEM+) for client-centered and participation-focused care planning. Am J Occup Ther 2019 Jun 10;73(4):1-8. [doi: 10.5014/ajot.2019.032235] [Medline: 31318677]
141. Jarvis JM, Kaelin VC, Anaby D, Teplicky R, Khetani MA. Electronic participation-focused care planning support for families: a pilot study. Dev Med Child Neurol 2020 Aug 07;62(8):954-961 [FREE Full text] [doi: 10.1111/dmcn.14535] [Medline: 32259292]
142. Khetani MA, Lim HK, Corden ME. Caregiver input to optimize the design of a pediatric care planning guide for rehabilitation: descriptive study. JMIR Rehabil Assist Technol 2017 Oct 24;4(2):e10. [doi: 10.2196/rehab.7566]

143. Khetani MA, Cliff AB, Schelly C, Daunhauer L, Anaby D. Decisional support algorithm for collaborative care planning using the participation and environment measure for children and youth (PEM-CY): a mixed methods study. Phys Occup Ther Pediatr 2014 Mar 26;35(3):231-252. [doi: 10.3109/01942638.2014.899288] [Medline: 24670061]
144. Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. Artif Intell Med 2013 Jan;57(1):9-19. [doi: 10.1016/j.artmed.2012.12.003] [Medline: 23287490]
145. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 2020 Jul 31;11(1):3852 [FREE Full text] [doi: 10.1038/s41467-020-17431-x] [Medline: 32737308]
146. Farzana S, Valizadeh M, Parde N. Modeling dialogue in conversational cognitive health screening interviews. In: Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020). 2020 Presented at: 12th International Conference on Language Resources and Evaluation (LREC 2020); May 13-15, 2020; Marseille, France p. 1167-1177 URL: https://tinyurl.com/mdpeccs5
147. Farzana S, Parde N. Exploring MMSE score prediction using verbal and non-verbal cues. In: Procedings of the Interspeech 2020 Conference. 2020 Presented at: Interspeech 2020 Conference; Oct 25–29, 2020; Shanghai, China p. 2207-2211. [doi: 10.21437/Interspeech.2020-3085]
148. Di Palo F, Parde N. Enriching neural models with targeted features for dementia detection. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. 2019 Presented at: 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop; Jul 2019; Florence, Italy p. 1-7. [doi: 10.18653/v1/p19-2042]
149. Bosak DL, Jarvis JM, Khetani MA. Caregiver creation of participation-focused care plans using Participation and Environment Measure Plus (PEM+), an electronic health tool for family-centred care. Child Care Health Dev 2019 Nov 30;45(6):791-798. [doi: 10.1111/cch.12709] [Medline: 31313843]
150. Arakelyan S, Maciver D, Rush R, O'hare A, Forsyth K. Family factors associated with participation of children with disabilities: a systematic review. Dev Med Child Neurol 2019 May 06;61(5):514-522 [FREE Full text] [doi: 10.1111/dmcn.14133] [Medline: 30613957]
151. Bult MK, Verschuren O, Jongmans MJ, Lindeman E, Ketelaar M. What influences participation in leisure activities of children and youth with physical disabilities? A systematic review. Res Dev Disabil 2011 Sep;32(5):1521-1529. [doi: 10.1016/j.ridd.2011.01.045] [Medline: 21388783]
152. Shikako-Thomas K, Majnemer A, Law M, Lach L. Determinants of participation in leisure activities in children and youth with cerebral palsy: systematic review. Phys Occup Ther Pediatr 2008 May 10;28(2):155-169. [doi: 10.1080/01942630802031834] [Medline: 18846895]
153. More than numbers: a guide toward diversity, equity and inclusion (DEI) in data collection. Charles and Lynn Schusterman Family Foundation. 2020. URL: https://philanthropynewyork.org/resources/more-numbers-guide-toward-diversity-equity-and-inclusion-dei-data-collection-charles-and [accessed 2021-09-20]
154. Imms C, Adair B, Keen D, Ullenhag A, Rosenbaum P, Granlund M. 'Participation': a systematic review of language, definitions, and constructs used in intervention research with children with disabilities. Dev Med Child Neurol 2016 Jan 28;58(1):29-38 [FREE Full text] [doi: 10.1111/dmcn.12932] [Medline: 26411643]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
HMI: human-machine interaction
ML: machine learning
NLP: natural language processing
VR: virtual reality

Edited by G Eysenbach; submitted 13.11.20; peer-reviewed by AV Das, P Tripathi, C Yu, M Majurul Ahsan; comments to author 27.01.21; revised version received 21.05.21; accepted 07.09.21; published 04.11.21.

Please cite as:
Kaelin VC, Valizadeh M, Salgado Z, Parde N, Khetani MA
Artificial Intelligence in Rehabilitation Targeting the Participation of Children and Youth With Disabilities: Scoping Review
J Med Internet Res 2021;23(11):e25745
URL: https://www.jmir.org/2021/11/e25745
doi:10.2196/25745
PMID:34734833

Original Paper

Using Artificial Intelligence With Natural Language Processing to Combine Electronic Health Record's Structured and Free Text Data to Identify Nonvalvular Atrial Fibrillation to Decrease Strokes and Death: Evaluation and Case-Control Study

Peter L Elkin[1,2,3], MD; Sarah Mullin[1], PHO; Jack Mardekian[4], PhD; Christopher Crowner[1], MSc; Sylvester Sakilay[1], MSc; Shyamashree Sinha[1], MSc, MD; Gary Brady[4], DPH; Marcia Wright[4], PharmD; Kimberly Nolen[4], PharmD; JoAnn Trainer[4], PharmD; Ross Koppel[1], PhD; Daniel Schlegel[1], PhD; Sashank Kaushik[1], MD; Jane Zhao[1], MD; Buer Song[1], MD, PhD; Edwin Anand[1], MD

[1]Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, United States

[2]Bioinformatics Laboratory, Department of Veterans Affairs, VA Western New York Healthcare System, Buffalo, NY, United States

[3]School of Engineering, University of Southern Denmark, Odense, Denmark

[4]Pfizer, Inc., New York, NY, United States

Corresponding Author:
Peter L Elkin, MD
Department of Biomedical Informatics
University at Buffalo
77 Goodell St
Suite 5t40
Buffalo, NY, 14203
United States
Phone: 1 5073581341
Email: elkinp@buffalo.edu

Abstract

Background: Nonvalvular atrial fibrillation (NVAF) affects almost 6 million Americans and is a major contributor to stroke but is significantly undiagnosed and undertreated despite explicit guidelines for oral anticoagulation.

Objective: The aim of this study is to investigate whether the use of semisupervised natural language processing (NLP) of electronic health record's (EHR) free-text information combined with structured EHR data improves NVAF discovery and treatment and perhaps offers a method to prevent thousands of deaths and save billions of dollars.

Methods: We abstracted 96,681 participants from the University of Buffalo faculty practice's EHR. NLP was used to index the notes and compare the ability to identify NVAF, congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category (CHA_2DS_2-VASc), and Hypertension, Abnormal liver/renal function, Stroke history, Bleeding history or predisposition, Labile INR, Elderly, Drug/alcohol usage (HAS-BLED) scores using unstructured data (International Classification of Diseases codes) versus structured and unstructured data from clinical notes. In addition, we analyzed data from 63,296,120 participants in the Optum and Truven databases to determine the NVAF frequency, rates of CHA_2DS_2-VASc ≥2, and no contraindications to oral anticoagulants, rates of stroke and death in the untreated population, and first year's costs after stroke.

Results: The structured-plus-unstructured method would have identified 3,976,056 additional true NVAF cases (P<.001) and improved sensitivity for CHA_2DS_2-VASc and HAS-BLED scores compared with the structured data alone (P=.002 and P<.001, respectively), causing a 32.1% improvement. For the United States, this method would prevent an estimated 176,537 strokes, save 10,575 lives, and save >US $13.5 billion.

Conclusions: Artificial intelligence–informed bio-surveillance combining NLP of free-text information with structured EHR data improves data completeness, prevents thousands of strokes, and saves lives and funds. This method is applicable to many disorders with profound public health consequences.

(J Med Internet Res 2021;23(11):e28946) doi:10.2196/28946

KEYWORDS

afib; atrial fibrillation; artificial intelligence; NVAF; natural language processing; stroke risk; bleed risk; CHA2DS2-VASc; HAS-BLED; bio-surveillance

Introduction

Background

Atrial fibrillation (AF), the most common type of arrhythmia [1,2], consists of nonvalvular AF (NVAF) and valvular AF (VAF) [1]. NVAF comprises approximately 70% of AF and currently affects approximately 5.8 million US patients and approximately 11 million in Europe on VAF results in a five times greater risk of stroke [3] and causes approximately 15% of all strokes [2,4]. Anticoagulation treatment dramatically reduces one's odds of a stroke to <0.5% on average.

The incidence of stroke with AF has prompted the development of scoring risk systems to guide anticoagulation treatment [5,6]. In 2014, the American Heart Association, American College of Cardiology, and Heart Rhythm Society advocated for AF practice guidelines via the use of congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category (CHA_2DS_2-VASc) scores that combine the $CHADS_2$ score with additional moderate risk factors [2,7]. Individuals' stroke risks should inform therapeutic options, which may include anticoagulants [7]. The Hypertension, Abnormal liver/renal function, Stroke history, Bleeding history or predisposition, Labile INR, Elderly, Drug/alcohol usage (HAS-BLED) score is a practical tool to assess individuals' risk of major bleeding and to guide anticoagulant therapy [8,9]. Researchers posit that the assessment of bleeding risk factors—age, uncontrolled hypertension, ischemic heart disease, and prior ischemic stroke—may improve individualized treatment for AF.

However, despite strong recommendations, oral anticoagulation (OAC) for NVAF patients remains low, with rates ranging from 39%-65% [10]. Disease surveillance and clinical decision support could help detect potential candidates who could benefit from this therapy. Automatic extraction from electronic health records (EHRs) has been shown to aid health care providers by making health care information easily accessible and helping with risk calculation [11,12]. Using these tools could reduce clinicians' computer time for data retrieval and data entry and could facilitate capturing all qualifying patients [13].

The Need for Natural Language Processing

Although EHRs contain an abundance of codified information, factors related to the assessment of NVAF are often poorly reflected in structured data [11]. Clinical text harboring rich contextual medical information is unstructured and in free-text form. Extracting information from a clinical text remains challenging because of context-specific abbreviations, refusal to adhere to typical language conventions, and because text often includes a broad range of specific medical terms. To retrieve information from a clinical text, multiple natural language processing (NLP) approaches have been developed, including those that extract clinical entities and map them to clinical terminologies such as SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) [14].

To capture all potential patients with NVAF and of CHA_2DS_2-VASc >1 who would benefit from appropriate anticoagulation therapy, we developed a method to automate risk scoring systems using a combination of multiple EHR data sources for diagnostic information, namely the International Classification of Disease (ICD) codes and clinical notes and lists. As natural language processors are expensive to develop and require individual tuning for each task or disease area, we make use of a high definition-NLP (HD-NLP) method that uses semisupervised learning to surpass the classification performance that could be obtained either by discarding the unlabeled data and performing supervised learning or by discarding the labels and performing unsupervised learning [15]. We compare the advantages of using NLP tools for NVAF phenotyping and calculate the risk scores of using structured ICD data alone.

Methods

This study compares the effectiveness of identifying NVAF patients using three methods: (1) structured EHR data, (2) a combination of structured EHR data and NLP-analyzed existing free text (EHR notes, problem lists, and laboratories), and (3) clinicians' assessments of NVAF patients (*the gold standard*). We used NLP of the EHRs' free text to improve the identification of NVAF patients and to assess their stroke and bleeding risks more accurately. We verified the improvement in the identification of NVAF cases and in determining the CHA_2DS_2-VASc and HAS-BLED scores. We then examined the rates of NVAF and treatment in patients with a CHA_2DS_2-VASc of ≥2 and no contraindications to treatment to determine the results from our local population. Finally, we extrapolated our findings on NVAF numbers to the US population and disease costs.

Study Populations

We had two samples: a local Western New York population of 96,681 individuals and 63,296,120 participants from the Optum and Truven databases.

Sample 1: Local

To understand the effectiveness of the system in identifying NVAF patients who should be treated and are not currently on OAC therapy, we abstracted a set of 96,681 participants (aged 18-90 years) from the Allscripts outpatient electronic records at the University at Buffalo's (UBMD) faculty practice. The research was approved by the institutional review board of the University of Buffalo.

Patient data were abstracted from 2010 to September 21, 2015, before the switch to ICD-10, allowing consistent use of ICD-9 terminology and sufficient follow-up data for the study period. This yielded 212,343 patients. Of those 212,343 patients, 96,681 (45.53%) had notes and were seen for ≥1 outpatient visits (Multimedia Appendix 1, Figure S1). Outcomes from these data included rates of AF, NVAF, and VAF diagnosis, components of the CHA_2DS_2-VASc and HAS-BLED scores, relevant contraindications, OAC treatment, and demographic variables. We excluded patients if they were on oral antithrombotic therapy for indications other than NVAF, had a mechanical prosthetic valve, had a hemodynamically significant mitral stenosis or significant aortic stenosis, were pregnant, had a transient AF because of reversible conditions, or had active infective endocarditis (Multimedia Appendix 1, Figure S2). We developed the NVAF cohort using ICD-9 codes (structured data) and ICD-9 and NLP (structured-plus-unstructured) of EHR notes and patient problems. AF and atrial flutter were defined by ICD-9 codes 427.31 and 427.32 and by SNOMED CT codes 49436004 and 5370000 with all subtypes in the hierarchy.

The structured data–only method used ICD 9 codes from problem lists, medications, and demographics. The structured-plus-unstructured method added clinical notes, vital signs, laboratory findings, and text from the problem list using HD-NLP for codification [14,16-18]. Free text elements were coded using SNOMED CT, a general description logic–based nomenclature of clinical medicine. Specific code inclusions can be found in Multimedia Appendix 1, Figure S3.

We then compared the accuracy of structured data alone with the structured-plus-unstructured EHR data derived using the HD-NLP system, focusing on the two models' abilities to identify true cases of NVAF and to determine stroke and bleeding risks (CHA_2DS_2-VASc and HAS-BLED scores).

Subsample of the Local Data

For validation of the accuracy of NLP, we used a gold standard created by human review (BS, JZ, EA, and SS) from a random sample of 300 patients. To verify the NVAF identification and CHA_2DS_2-VASc and HAS-BLED scores, we used this 300-patient random sample from our NVAF patients, which were dual human reviewed. We also looked to determine how much better structured-plus-unstructured data were in the identification of NVAF cases and in the determination of the CHA_2DS_2-VASc and HAS-BLED scores.

The human review data set was independently examined by 4 clinicians, each performing 150 reviews on deidentified patient encounters from the EHR. Each clinician made a judgment as to whether the patients had sustained NVAF and whether the patient had each of the components of the CHA_2DS_2-VASc and HAS-BLED scores. If there were disagreements, a fifth clinician adjudicated.

Calculations determined that 300 patients were needed for 90% power to predict a 5% change in accuracy given a two-sided alpha of .05, assuming a standard accuracy of 73% based on ICD-9 codes [19]. Multimedia Appendix 1, Figure S1 presents the decision tree and sample numbers, and Multimedia Appendix 1, Figure S2 illustrates the randomization scheme.

Sample 2: National—Optum and Truven Databases

We analyzed the claims data from 63,296,120 participants in the Optum and Truven databases from October 2015 to September 2016 to determine the frequency of NVAF, rates of CHA_2DS_2-VASc ≥2, and no contraindications to OAC, rates of stroke and death in the untreated NVAF, strokes and death in the large claims database, and the first year's cost after stroke [20,21]. Cost differences were based on 1-year cost before and after the stroke, adjusted for inflation.

We then extrapolated our findings to the US population.

Findings for NLP

We made use of an HD-NLP to rapidly assign ontological terms to the text in patient records (Multimedia Appendix 1, Figure S5) [14,16,17]. HD-NLP is a full-function NLP processing pipeline that takes sentences, parses them by their parts of speech, and builds a full semantic parse in memory; then, an ontological coder works by matching words to ontology terms, with the longest match being preferred. We used basic formal ontology as an upper-level ontology to index the data from individual trials [18]. We also used the ontology of biomedical investigation and SNOMED CT as our main ontologies [22,23].

A level of syntactic processing was required to match text with ontological terms. The linguistic representation is specified in language models. Of primary concern here was an English language model to identify sentences, phrases, words, and parts of speech. Terms from the input ontologies were then assigned to spans of text. String matching techniques allowed for inexact matches influenced by the underlying language model. The structures of the free-text medical records were captured and stored.

To develop the NVAF model, we used a semisupervised learning algorithm training set with 36,268 patients from the Allscripts EHR UBMD faculty practice data from 2007 to 2008, with 1972 AF cases and 1795 NVAF cases to determine the best SNOMED CT codes to match the case definition. As most clinical texts are unlabeled, semisupervised learning leverages a small amount of labeled data with a large amount of unlabeled data. Researchers have shown that large amounts of unlabeled data, when used in conjunction with a limited amount of labeled data, can produce considerable improvement in learning accuracy, especially with assistance from subject matter expert's annotation of the training set's false positive and false negative results from each training iteration [14]. All cases were coded using HD-NLP with SNOMED CT codes (the unsupervised portion of the study). Where the SNOMED CT codes and ICD-9 codes agreed that the patient had NVAF, we called that a true positive case. The same logic was used to determine true negatives. Where either coding system disagreed, our clinician (PE) reviewed the case and decided. After reviewing the false positive and false negative cases from the training data set, we used additional synonymy to the terminology and selected a more appropriate set of codes for each rule in the definition. This process was iterated on the training set until we met our accuracy goals.

Statistical Analysis

Statistical analyses were conducted using R 3.3.2. A random gold standard sample of 300 patients was taken from the sample 1 AF cohort defined by both ICD and HD-NLP. Interrater agreement was assessed using the two-way random effects model for intraclass correlation coefficient, with two-sided 10,000 samples bootstrapped 95% CI, treating the risk scores as continuous. Cohen κ with two-sided 10,000 samples bootstrapped 95% CI assessed the interrater reliability of each individual component of the scores, NVAF and AF.

The accuracy of the structured data alone was compared with structured-plus-unstructured data for the outcomes of NVAF, CHA_2DS_2-VASc score, and HAS-BLED score in the random sample. Cohen κ with two-sided bootstrapped CIs was calculated as a measure of reliability between the gold standard and the structured and structured-plus-unstructured data. For sensitivity and specificity, a hypothesis test comparing structured with structured-plus-unstructured data was assessed using either the McNemar test for paired observations or the binomial exact test. For positive and negative predictive values, a generalized score statistic proposed by Leisenring et al [19] was used for comparison.

As the CHA_2DS_2-VASc and HAS-BLED scores are on ordinal scales from 0 to 9, we analyzed the area under the receiver operator characteristic curve (ROC) using the C-Index and Somer D, based on ordinal logistic regression, where probabilities were modelled as $P(Y \geq k|X)$, where k defines the cut-offs from 0 to 9 that the score can take. We hypothesized that the structured and NLP data were more concordant than the structured-only data compared with the gold standard between the ordinal gold standard score and the ordinal method score.

We contrasted our findings with the clinical judgments from the physician review of the 300 patients, categorized as contraindicated (Multimedia Appendix 1, Table S1) or not on OAC, would or would not benefit from OAC, and not on OAC. To determine the potential effects of adopting the NLP-enabled method with structured-plus-unstructured data, the accuracy data of the structured and NLP data method were used to extrapolate the findings for all untreated US patients in the Optum and Truven data sets with no contraindications to OACs. Then, the potential savings from reduced strokes were derived and compared with the prevailing structured-only method.

Results

NLP Results

From the Allscripts UBMD practice EHR data, we found 2722 potential patients with NVAF using the structured and NLP method and 1849 cases using only ICD-9 codes. The use of NLP by combining structured-plus-unstructured data improved sensitivity by 32.1%, that is, 873/2722 (P<.001) in determining the NVAF population. In the random sample, participants were on average 72 years old (mean 72.7, SD 13.6), 41.3% (125/300) were female, and 86.3% (259/300) were White. The true NVAF population within the random sample, as determined by clinician review, was 88% (264/300) of cases with an average age of 73 (mean 73.4, SD 13.0), of which 41.7% (110/264) were female, and 87.1% (230/264) were White. The assessment of agreement between clinicians and interrater reliability was high for the CHA_2DS_2-VASc score (odds ratio [OR] 0.796, 95% CI 0.725-0.853 and OR 0.878, 95% CI 0.838-0.909) and adequate for the HAS-BLED score (OR 0.609, 95% CI 0.51-0.692 and OR 0.675, 95% CI 0.544-0.77). Cohen κ, depending on whether an outcome was a rare event, ranged from –0.080 to 0.84.

When we tested this in the human review of the 300 cases, we found a 46% improvement in sensitivity (Table 1), which is greater than the 32.1% improvement seen with the automated method.

Table 1. Clinician review (gold standard): comparison of outcomes for structured and structured-plus-unstructured data against the gold standard for identifying a case as nonvalvular atrial fibrillation.

Outcome	Structured surveillance	Structured and NLP[a] surveillance	P value
Sensitivity, OR[b] (95% CI)	0.54 (0.48-0.60)	1 (0.979-1)	<.001
PPV[c], OR (95% CI)	0.95 (0.90-0.98)	0.93 (0.893-0.956)	.24
F[d] score	0.686	0.964	N/A[e]

[a]NLP: natural language processing.

[b]OR: odds ratio.

[c]PPV: positive predictive value.

[d]For case finding of nonvalvular atrial fibrillation.

[e]N/A: not applicable.

Thus, the structured-plus-unstructured surveillance showed that the sensitivity for CHA_2DS_2-VASc ≥2 and HAS-BLED≥3 scores was significantly better than that for structured data alone (P=.002 and P<.001, respectively). The specificities of the two methods were not statistically different for CHA_2DS_2-VASc and favored the structured method for HAS-BLED (Table 2). The positive predictive value (PPV; precision) also improved for the HAS-BLED score using the structured-plus-unstructured method (Table 2) but was not statistically different from the structured data for the CHA_2DS_2-VASc score. However, the negative predictive value improved for both scores using the structured-plus-unstructured method. No cases identified by the structured method were missed by the structured-plus-unstructured method.

Table 2. Comparison of outcomes for structured and structured-plus-unstructured surveillance against the clinician review (gold standard) for identifying Hypertension, Abnormal liver/renal function, Stroke history, Bleeding history or predisposition, Labile INR, Elderly, Drug/alcohol usage (HAS-BLED) and congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category (CHA2DS2-VASc) components.

Method	HAS-BLED					CHA_2DS_2-VASc				
	Structured surveillance	Structured and NLP[a] surveillance	Difference	*T* test	*P* value	Structured surveil-lance	Structured and NLP surveillance	Difference	Test statis-tic	*P* value
Sensitivity										
McNemar method	0.382	0.806	0.424	72	<.001	__[b]	—	—	—	—
Exact binomial method	—	—	—	—		0.942	0.983	0.0413	—	.002
Specificity										
McNemar method	0.947	0.777	–0.17	16	<.001					
Exact binomial method	—	—	—	—	—	0.955	0.909	–0.0455		>.99[c]
PPV[d]										
Generalized score method	0.929	0.867	.061	4.487	.03	0.996	0.992	0.004	0.915	.34
NPV[e]										
Generalized score method	0.459	0.689	0.23	47.757	<.001	0.6	0.833	0.233	11.662	<.001

[a]NLP: natural language processing.

[b]There is a small number of discordant cells, such that for the gold standard's CHA_2DS_2-VASc <2, there is 1 case that was identified as CHA_2DS_2-VASc ≥2 in the structured and NLP method but not in the structured method. The exact binomial *P* value is calculated as

[c]There is a small number of discordant cells, such that for the gold standard's CHA_2DS_2-VASc <2, there is 1 case that was identified as CHA_2DS_2-VASc >2 in the Structured and NLP method but not in the structured method. The exact binomial *P* value is calculated as

[d]PPV: positive predictive value.

[e]NPV: negative predictive value.

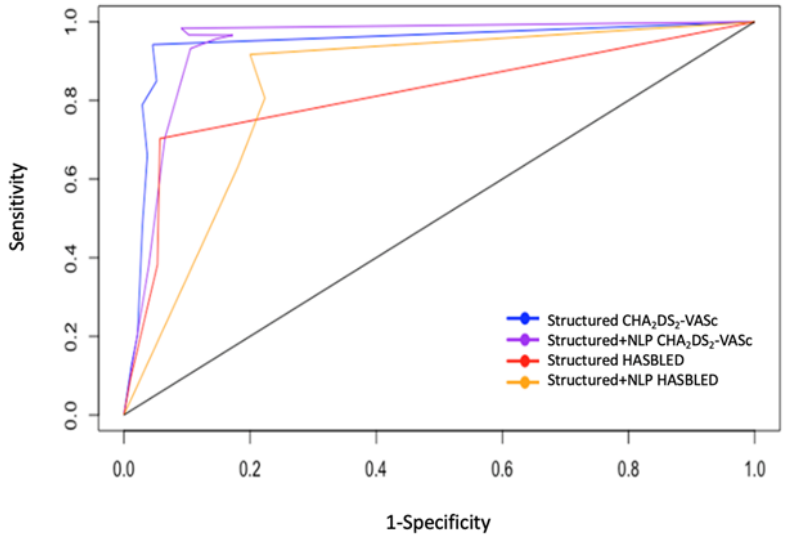
Multimedia Appendix 1, Figure S4 presents the conditional probability tree for the automated structured or structured-plus-NLP method, based on clinical guidelines.

In Figure 1, the area under the ROC for the CHA_2DS_2-VASc scores for the structured-plus-unstructured data compared with the gold standard score was 0.914 (95% CI 0.896-0.933) with a Somer D 0.829 (SD 0.0185), and for the structured data alone compared with the gold standard score, was 0.863 (CI 0.838-0.887), with a Somer D 0.726 (SD 0.0249). For CHA_2DS_2-VASc scores, structured-plus-unstructured data were more concordant than structured data alone when compared with the gold standard score (Z=19.77; *P*<.001). For the ROC curves of the HAS-BLED scores with the gold standard score as the outcome, the structured-plus-unstructured data was 0.816 (CI 0.783-0.849), with a Somer D 0.633 (SD 0.034), and the structured data alone was 0.797 (CI 0.761-0.833) with a Somer D 0.595 (SD 0.037). For HAS-BLED scores, structured-plus-unstructured data were not more concordant than structured data alone (Z=1.433; *P*=.149).

Figure 1 represents four areas under ROC curves, two for structured versus structured and NLP CHA_2DS_2-VASc score and two for structured versus structured and NLP HAS-BLED score. As these scores are ordinal (eg, ranging from 0-9) and not binary, as with typical ROC, we use the C-Index and Somer D based on ordinal logistic regression to model the probabilities, resulting in multiple y values for the same x.

We compared the findings of the gold standard with the NLP structured-plus-unstructured data (Multimedia Appendix 1, Table S1). Clinician reviewers found 31 untreated patients who should have been treated and 1 treated patient who, the clinicians felt, should not have been treated. This was the same total as that of the gold standard. After clinician review, there was a 32.1% improvement in PPV using the structured-plus-unstructured method when compared with the structured method alone.

Figure 1. Four receiver operator characteristic curves for cumulative congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category (CHA2DS2-VASc), and Hypertension, Abnormal liver/renal function, Stroke history, Bleeding history or predisposition, Labile INR, Elderly, Drug/alcohol usage (HAS-BLED) risk scores. NLP: natural language processing.



Extrapolating Findings to the US Population for Prevalence and Cost

Extrapolation to the US population of the Truman and Optum data results can be found in Table 3.

To determine the national cost savings from the NLP-assisted bio-surveillance of the structured-plus-unstructured data, we used Truven data and contrasted the mean monthly costs per patient after a stroke (US $11,538) with the monthly costs before a stroke (US $2,763.33), which yielded a mean savings of US $8,776.02. This was adjusted to US $2019 as the data were from 2010 to 2015. This revealed savings of US $8,556.66 per month or yearly savings of US $102,680.

The structured data method identified 1.5% (967,801/63,296,120) of the population as having NVAF. Of those cases, 84.3% (816,240/967,801) had a CHA_2DS_2-VASc score of ≥2. These data indicate that 60.7% (495,749/816,240) of these patients were not treated despite the current clinical guidelines. Untreated NVAF patients had a 4.4% (22,021/495,749) annual ischemic stroke risk, and the stroke patients had a 6.0% (1320/22,021) risk of death.

Table 3. Optum and Truven stroke data for 1 year after atrial fibrillation (AF) diagnosis.

Population for rates	Truven, n (%)	Optum, n (%)	Total, n (%)	Event rates (%)
All patients	32,046,193 (50.63)	31,249,927 (49.37)	63,296,120 (100)	—[a]
Patients aged ≥18 years in 2016 with any diagnosis of AF during October 2015-September 2016	422,092 (32.79)	865,072 (67.21)	1,287,164 (100)	—
Patients aged ≥18 years in 2016 with any diagnosis of AF during October 2015-September 2016 and without a VHD[b] diagnosis during 1-year preindex	355,811 (36.76)	611,990 (63.24)	967,801 (100)	1.5
Patients aged ≥18 years in 2016 with any diagnosis of AF during October 2015-September 2016 and without VHD diagnosis during 1-year preindex and with CHA_2DS_2-VASc[c] ≥2 and no contraindications to OAC[d]	276,465 (33.87)	539,775 (66.13)	816,240 (100)	84.3
Patients aged ≥18 years in 2016 with any diagnosis of AF during October 2015-September 2016 and without VHD diagnosis during 1-year preindex and with CHA_2DS_2-VASc ≥2 and no contraindications to OAC and were untreated	179,441 (36.20)	316,308 (63.80)	495,749 (100)	60.7
Stroke rate	11,530 (52.36)	10491 (47.64)	22,021 (100)	4.4
Death rate	727 (55.1)	593 (44.9)	1,320 (100)	5.99

[a]The values are not events.

[b]VHD: valvular hear disease.

[c]CHA_2DS_2-VASc: congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category.

[d]OAC: oral anticoagulation.

Estimates of Morbidity, Mortality, and Cost

After extrapolating our results combining the Optum and Truven data with our method of bio-surveillance, we estimated outcomes of implementing the NLP-assisted analyses of structured-plus-unstructured data nationally; that is, if implemented nationally (among a population of 316,005,000), this system could potentially prevent 176,537 strokes and 10,575 deaths in the first year of implementation, with stroke-associated savings >US $18.126 billion (Table 4).

Table 4. Untreated strokes and their costs for first year after the event.

Extrapolated results	Structured surveillance	Structured and NLP[a] surveillance	Difference between the two methods
NVAF[b] population	4,955,284	6,545,930	1590,646
NVAF population with no contraindications and CHA_2DS_2VASc[c] ≥2	4,543,995	6,002,707	1,458,712
NVAF population needing treatment	3,009,840	3,976,057	966,217
Strokes prevented	133,637	176,537	42,900
Deaths prevented	8,005	10,575	2,570
Cost savings[d] (US $)	13,721,820,000	18,126,800,000	4,404,981,210

[a]NLP: natural language processing.

[b]NVAF: nonvalvular atrial fibrillation.

[c]CHA_2DS_2-VASc: congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category.

[d]Cost basis is US $102,680 per untreated ischemic stroke patient's excess cost for the first year after event; cost is 1.9% inflation adjusted.

Discussion

Principal Findings

Compared with structured EHR data alone, we found that NLP-assisted structured-plus-unstructured EHR data identified previously unknown and untreated patients with NVAF and their stroke and bleed risks with greater accuracy. Adding the unstructured data significantly improved the sensitivity and negative predictive value across all measures, whereas the results for NVAF specificity and PPV were strong but mixed. Future applications of this artificial intelligence (AI) bio-surveillance method may involve identifying other underdiagnosed populations.

We estimated NVAF rates in large national database populations, the percentage of people who should be treated with OAC and are not currently treated, and yearly risks of stroke expressed as a percentage of these untreated patients [24,25]. We also estimated the average incremental 1-year cost for a stroke event and identified stroke-related average death rates in the first year after event.

Verhoef et al [26,27] showed that bleeding rates with warfarin were, on average, 0.34% risk per year. Given additional treatment for 3,976,057 new patients, we would expect 13,824 new patient bleeds. McWilliam [28] showed that the average cost of a major bleed was US $19,000 in 2008 (inflation adjusted to US $23,777.67). For the population, this equals US $328,702,452. Gilligan et al [29] showed that the average total cost for warfarin therapy was US $76.19 per member per month, which translates to a total national cost of US $3,750,758,790 per year. Potential net financial treatment benefits from using the NLP-assisted structured-plus-unstructured method equates to US $14.4 billion (US $18.13 billion to US $3.75 billion).

On the basis of the accuracy of the AI-derived bio-surveillance method, we show potential societal benefits of implementing this technology. Nationally, this method could identify approximately 4 million patients requiring treatment, potentially preventing >176,000 strokes in the first year, and >10,500 deaths, translating to national savings of >US $14 billion. Including the estimated costs of excess bleeding from the treatment and from our estimate, the national implementation costs would be no greater than US $300,000,000. This type of AI-driven clinical decision support bio-surveillance has the potential to significantly improve patient care and clinicians' treatment decisions.

NVAF is but one important condition among many. Future applications of this AI bio-surveillance method may identify other underdiagnosed populations. Once deployed, the infrastructure could be used for other disorders and could be implemented at a low incremental cost.

Limitations

This analysis and data extrapolation were based on previous 2014 American Heart Association, American College of Cardiology, and Heart Rhythm Society recommendations for OAC therapy in patients with NVAF and a CHA_2DS_2-VASc score of ≥2. The 2019 focused updates on AF now recommend that men with a CHA_2DS_2-VASc score of ≥2 and women with a CHA_2DS_2-VASc score of ≥3 should be treated with an OAC. As such, the numbers in this analysis may include women who, under the updated guidance, may not be recommended for treatment with an OAC. In addition, not all patients for whom therapy is indicated may agree to accept anticoagulation therapy.

The Optum and Truven databases, although found to be effectively nonoverlapping, are, on average, considered to be for younger and healthier private payer populations; therefore,

we may underestimate both protective effects and cost savings [30]. If this method were extended to other diseases, models must be built and distributed uniformly across the country and perhaps internationally.

The AI model processes the free text of the notes and reports, and as it can accept and process data from Cerner, Epic, and other EHRs, there should be no difference in outcome; however, this model has not been specifically tested with data from other EHRs.

ICD-9 codes were used in this study because of the desire to have a consistently coded data set. ICD-10 codes were not included. Future research should investigate this method using later ICD codes.

This informatics method promises many benefits. Of course, additional research is needed to determine its applicability to other diseases.

Conclusions

Although a common disorder (N=6 million Americans), NVAF is often underprophylaxed for thromboembolic events that may lead to strokes. Critical evidence may be found in patients' EHRs to aid in anticoagulation decision-making. Stroke rates of untreated patients with a CHA_2DS_2-VASc of ≥2 in our study were 4.44%, and of these, approximately 6% will die within 1 year. Treatment dramatically reduces one's odds of a stroke to <0.5% on average.

Our structured-plus-unstructured (NLP) method identified 36.3% additional true NVAF cases (P<.001) compared with the structured data alone. Extrapolating to the US population using the 63 million people in the Optum and Truven populations allowed us to predict that in just the first-year implementation of this system, it could prevent 176,537 strokes and 10,575 deaths and save the nation >US $13.5 billion dollars.

Moreover, this bio-surveillance method and preparedness, in general, may be useful for the discovery and treatment of many other disorders, and require further research with different diseases. Automated tools in partnership with clinicians have the potential to significantly improve adherence to established clinical guidelines and to precision medicine.

Acknowledgments

This study was funded in part by Pfizer, Inc, and grants were received from the National Institutes of Health (NIH)-National Library of Medicine T15LM012495, National Center for the Advancement of Translational Sciences UL1TR001412, National Institute of Alcohol Abuse and Alcoholism R21AA026954, R33AA026954 and NIH T32GM099607. This study was funded in part by the National Cancer Institute and the Department of Veterans Affairs through the BD-STEP program. Optum's deidentified Integrated Claims-Clinical Data set and Truven Health MarketScan Research Databases were used in this study. Partial funding from Pfizer, NIH-NLM-T15LM012495, NCATS-UL1TR001412, NIH-T32GM099607, NCI, and US-VA was received (Multimedia Appendix 2).

Conflicts of Interest

GB, MW, JM, JT, and KM are employed at Pfizer.

Multimedia Appendix 1
Study recruitment diagram and additional analyses.
[DOCX File , 6156 KB - jmir_v23i11e28946_app1.docx]

Multimedia Appendix 2
Financial and nonfinancial support.
[DOCX File , 14 KB - jmir_v23i11e28946_app2.docx]

References

1. Camm AJ, Lip GY, De Caterina R, Savelieva I, Atar D, Hohnloser SH, ESC Committee for Practice Guidelines (CPG). 2012 focused update of the esc guidelines for the management of atrial fibrillation: an update of the 2010 esc guidelines for the management of atrial fibrillation. Developed with the special contribution of the european heart rhythm association. Eur Heart J 2012 Nov;33(21):2719-2747. [doi: 10.1093/eurheartj/ehs253] [Medline: 22922413]
2. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society. J Am Coll Cardiol 2014 Dec 02;64(21):1-76 [FREE Full text] [doi: 10.1016/j.jacc.2014.03.022] [Medline: 24685669]
3. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation 2012 Jan 3;125(1):2-220 [FREE Full text] [doi: 10.1161/CIR.0b013e31823ac046] [Medline: 22179539]

4. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the framingham study. Stroke 1991 Aug;22(8):983-988. [doi: 10.1161/01.str.22.8.983] [Medline: 1866765]
5. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the national registry of atrial fibrillation. J Am Med Assoc 2001 Jun 13;285(22):2864-2870. [doi: 10.1001/jama.285.22.2864] [Medline: 11401607]
6. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The Euro Heart Survey on atrial fibrillation. Chest 2010 Feb;137(2):263-272. [doi: 10.1378/chest.09-1584] [Medline: 19762550]
7. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart rhythm society in collaboration with the society of thoracic surgeons. Circulation 2019 Jul 09;140(2):125-151. [doi: 10.1161/CIR.0000000000000665] [Medline: 30686041]
8. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (has-bled) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the euro heart survey. Chest 2010 Nov;138(5):1093-1100. [doi: 10.1378/chest.10-0134] [Medline: 20299623]
9. Lip GY, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the has-bled (hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile inr, elderly, drugs/alcohol concomitantly) score. J Am Coll Cardiol 2011 Jan 11;57(2):173-180 [FREE Full text] [doi: 10.1016/j.jacc.2010.09.024] [Medline: 21111555]
10. Tiryaki F, Nutescu EA, Hennenfent JA, Karageanes AM, Koesterer LJ, Lambert BL, et al. Anticoagulation therapy for hospitalized patients: patterns of use, compliance with national guidelines, and performance on quality measures. Am J Health Syst Pharm 2011 Jul 01;68(13):1239-1244. [doi: 10.2146/ajhp100543] [Medline: 21690430]
11. Navar-Boggan AM, Rymer JA, Piccini JP, Shatila W, Ring L, Stafford JA, et al. Accuracy and validation of an automated electronic algorithm to identify patients with atrial fibrillation at risk for stroke. Am Heart J 2015 Jan;169(1):39-44. [doi: 10.1016/j.ahj.2014.09.014] [Medline: 25497246]
12. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]
13. Aakre C, Dziadzko M, Keegan MT, Herasevich V. Automating clinical score calculation within the electronic health record. A feasibility assessment. Appl Clin Inform 2017 Apr 12;8(2):369-380 [FREE Full text] [doi: 10.4338/ACI-2016-09-RA-0149] [Medline: 28401245]
14. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. Ann Intern Med 2012 Jan 03;156(1 Pt 1):11-18. [doi: 10.7326/0003-4819-156-1-201201030-00003] [Medline: 22213490]
15. Yeung S, Downing NL, Fei-Fei L, Milstein A. Bedside computer vision - moving artificial intelligence from driver assistance to patient safety. N Engl J Med 2018 Apr 05;378(14):1271-1273. [doi: 10.1056/NEJMp1716891] [Medline: 29617592]
16. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. J Am Med Assoc 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
17. Schlegel DR, Crowner C, Lehoullier F, Elkin PL. HTP-NLP: A new NLP system for high throughput phenotyping. Stud Health Technol Inform 2017;235:276-280 [FREE Full text] [Medline: 28423797]
18. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care 2005 May;43(5):480-485. [doi: 10.1097/01.mlr.0000160417.39497.a9] [Medline: 15838413]
19. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics 2000 Jun;56(2):345-351. [doi: 10.1111/j.0006-341x.2000.00345.x] [Medline: 10877288]
20. Optum. 2019. URL: https://www.optum.com/solutions/data-analytics/data.html [accessed 2019-02-02]
21. Truven Health Analytics. IBM. 2019. URL: https://truvenhealth.com/Portals/0/assets/ACRS_11223_0912_MarketScanResearch_SS_Web.pdf [accessed 2919-02-02]
22. Wade G. Implementing snomed ct for quality reporting: avoiding pitfalls. Appl Clin Inform 2011;2(4):534-545 [FREE Full text] [doi: 10.4338/ACI-2011-10-RA-0056] [Medline: 23616894]
23. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The ontology for biomedical investigations. PLoS One 2016;11(4):e0154556 [FREE Full text] [doi: 10.1371/journal.pone.0154556] [Medline: 27128319]
24. Bray BD, Paley L, Hoffman A, James M, Gompertz P, Wolfe CD, et al. Socioeconomic disparities in first stroke incidence, quality of care, and survival: a nationwide registry-based cohort study of 44 million adults in england. Lancet Public Health 2018 Apr;3(4):185-193 [FREE Full text] [doi: 10.1016/S2468-2667(18)30030-6] [Medline: 29550372]
25. Freedman B, Potpara TS, Lip GY. Stroke prevention in atrial fibrillation. Lancet 2016 Aug 20;388(10046):806-817. [doi: 10.1016/S0140-6736(16)31257-0] [Medline: 27560276]

26. Verhoef TI, Redekop WK, Darba J, Geitona M, Hughes DA, Siebert U, EU-PACT Group. A systematic review of cost-effectiveness analyses of pharmacogenetic-guided dosing in treatment with coumarin derivatives. Pharmacogenomics 2010 Jul;11(7):989-1002. [doi: 10.2217/pgs.10.74] [Medline: 20602617]
27. Yao X, Abraham NS, Sangaralingham LR, Bellolio MF, McBane RD, Shah ND, et al. Effectiveness and safety of dabigatran, rivaroxaban, and apixaban versus warfarin in nonvalvular atrial fibrillation. J Am Heart Assoc 2016 Jun 13;5(6):e003725 [FREE Full text] [doi: 10.1161/JAHA.116.003725] [Medline: 27412905]
28. Desai JR, Hyde CL, Kabadi S, St Louis M, Bonato V, Loomis AK, et al. Utilization of positive and negative controls to examine comorbid associations in observational database studies. Med Care 2017 Mar;55(3):244-251 [FREE Full text] [doi: 10.1097/MLR.0000000000000640] [Medline: 27787351]
29. McWilliam A, Lutter R, Nardinelli C. Healthcare impact of personalized medicine using genetic testing: an exploratory analysis for warfarin. Per Med 2008 May;5(3):279-284. [doi: 10.2217/17410541.5.3.279] [Medline: 29783488]
30. Gilligan AM, Gandhi P, Song X, Wang C, Henriques C, Sander S, et al. All-cause, stroke-, and bleed-specific healthcare costs: comparison among patients with Non-Valvular Atrial Fibrillation (NVAF)newly treated with dabigatran or warfarin. Am J Cardiovasc Drugs 2017 Dec;17(6):481-492 [FREE Full text] [doi: 10.1007/s40256-017-0244-1] [Medline: 28795348]

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
CHA2DS2-VASc: congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category
EHR: electronic health record
HAS-BLED: Hypertension, Abnormal liver/renal function, Stroke history, Bleeding history or predisposition, Labile INR, Elderly, Drug/alcohol usage
HD-NLP: high definition natural language processing
ICD: International Classification of Disease
NIH: National Institutes of Health
NLP: natural language processing
NVAF: nonvalvular atrial fibrillation
OAC: oral coagulation
OR: odds ratio
PPV: positive predictive value
ROC: receiver operator characteristic curve
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
VAF: valvular atrial fibrillation

Edited by R Kukafka; submitted 19.03.21; peer-reviewed by H Oyama, P Sarajlic, S Bhatia; comments to author 15.04.21; revised version received 05.06.21; accepted 05.07.21; published 09.11.21.

Please cite as:
Elkin PL, Mullin S, Mardekian J, Crowner C, Sakilay S, Sinha S, Brady G, Wright M, Nolen K, Trainer J, Koppel R, Schlegel D, Kaushik S, Zhao J, Song B, Anand E
Using Artificial Intelligence With Natural Language Processing to Combine Electronic Health Record's Structured and Free Text Data to Identify Nonvalvular Atrial Fibrillation to Decrease Strokes and Death: Evaluation and Case-Control Study
J Med Internet Res 2021;23(11):e28946
URL: https://www.jmir.org/2021/11/e28946
doi:10.2196/28946
PMID:34751659

Original Paper

Predicting COVID-19–Related Health Care Resource Utilization Across a Statewide Patient Population: Model Development Study

Suranga N Kasturi[1,2], PhD; Jeremy Park[3], BS; David Wild[3], PhD; Babar Khan[1,4], MD, MS; David A Haggstrom[1,4], MD, MAS; Shaun Grannis[1,5], MD, MS

[1]Regenstrief Institute, Indianapolis, IN, United States

[2]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, United States

[3]Luddy School of Informatics Computing and Engineering, Indiana University, Bloomington, IN, United States

[4]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[5]Department of Family Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Suranga N Kasturi, PhD
Regenstrief Institute
1101 W 10th St
Indianapolis, IN, 46202
United States
Phone: 1 (317) 274 9000
Email: snkasthu@iu.edu

Abstract

Background: The COVID-19 pandemic has highlighted the inability of health systems to leverage existing system infrastructure in order to rapidly develop and apply broad analytical tools that could inform state- and national-level policymaking, as well as patient care delivery in hospital settings. The COVID-19 pandemic has also led to highlighted systemic disparities in health outcomes and access to care based on race or ethnicity, gender, income-level, and urban-rural divide. Although the United States seems to be recovering from the COVID-19 pandemic owing to widespread vaccination efforts and increased public awareness, there is an urgent need to address the aforementioned challenges.

Objective: This study aims to inform the feasibility of leveraging broad, statewide datasets for population health–driven decision-making by developing robust analytical models that predict COVID-19–related health care resource utilization across patients served by Indiana's statewide Health Information Exchange.

Methods: We leveraged comprehensive datasets obtained from the Indiana Network for Patient Care to train decision forest-based models that can predict patient-level need of health care resource utilization. To assess these models for potential biases, we tested model performance against subpopulations stratified by age, race or ethnicity, gender, and residence (urban vs rural).

Results: For model development, we identified a cohort of 96,026 patients from across 957 zip codes in Indiana, United States. We trained the decision models that predicted health care resource utilization by using approximately 100 of the most impactful features from a total of 1172 features created. Each model and stratified subpopulation under test reported precision scores >70%, accuracy and area under the receiver operating curve scores >80%, and sensitivity scores approximately >90%. We noted statistically significant variations in model performance across stratified subpopulations identified by age, race or ethnicity, gender, and residence (urban vs rural).

Conclusions: This study presents the possibility of developing decision models capable of predicting patient-level health care resource utilization across a broad, statewide region with considerable predictive performance. However, our models present statistically significant variations in performance across stratified subpopulations of interest. Further efforts are necessary to identify root causes of these biases and to rectify them.

(J Med Internet Res 2021;23(11):e31337) doi:10.2196/31337

KEYWORDS

COVID-19; machine learning; population health; health care utilization; health disparities; health information; epidemiology; public health; digital health; health data; pandemic; decision models; health informatics; healthcare resources

Introduction

Background

The COVID-19 pandemic has impacted the health and well-being of individuals, communities, and economies worldwide at an unprecedented scale [1,2]. As of June 1, 2021, the COVID-19 pandemic has infected over 170 million people worldwide and claimed the lives of over 3.5 million people. In the United States alone, COVID-19 has infected over 33 million people and claimed over 600,000 lives. In addition to the loss of lives and other adverse health outcomes, the enforcement of preventative measures, such as lockdowns and mask-wearing mandates, have further affected the mental and physical well-being of individuals and communities. The cumulative financial costs of the COVID-19 pandemic caused by lost output and health reduction has been estimated at US $16 trillion, or approximately 90% of the annual gross domestic product of the United States [3].

In the United States, the COVID-19 pandemic has highlighted (1) the inability of health systems to leverage existing system infrastructure in order to rapidly develop and apply broad analytical tools that could inform state- and national-level policymaking and patient care delivery in hospital settings and (2) systemic disparities in COVID-19–related outcomes and access to care based on race or ethnicity [4], gender [5], income level, and urban-rural divide [6,7]. At the peak of the pandemic outbreak in the United States, these limitations contributed to distrust, misinformation, and lack of cohesive decision-making. This impeded local government and public health officials from making informed policy decisions, such as mask-wearing mandates and stay-at-home orders, to control disease outbreaks and safeguard health systems from extended strain. This led to shortages in hospital beds, personal protective equipment, and health care personnel, thereby causing significant disruptions to health care delivery and consequent loss of lives [2,3].

Although the United States seems to be recovering from the COVID-19 pandemic owing to widespread vaccination efforts and increased public awareness, there is still a need to address the aforementioned limitations. Overcoming these limitations will ensure better disaster preparedness and response in anticipation of any future outbreaks caused by either COVID-19 variants or other diseases and to manage the care of vaccine-hesitant populations. The United States boasts significant health information system infrastructure, resulting in the active collection of a wide variety of patient-level clinical, medication, and visit history data. However, such datasets are often siloed across different health systems. As a result, analytical model development is often spearheaded at the health system level. Although such models may be useful in caring for a specific health system, they may not generalize across broader populations and cannot contribute to large-scale public health responses delivered across broad geographies, such as at the county, metropolitan area, or state level.

Objective

In this study, we sought to inform the feasibility of leveraging broad, statewide datasets for population health–driven decision-making by developing robust analytical models that predicted COVID-19–related health care resource utilization at the patient level among those served by Indiana's statewide Health Information Exchange (HIE).

Methods

Patient Population and Data Sources

We leveraged the COVID-19 Research Data Commons (CoRDaCo) [8], a rich, statewide dataset curated by the Regenstrief Institute of Indianapolis and Indiana University. The CoRDaCo dataset seeks to enable better access to data on COVID-19–positive patients for research purposes. It integrates data from multiple clinical sources, including the Indiana Network for Patient Care (INPC) [9]—one of the longest continuously operated statewide HIEs in the United States consisting of data from over 15 million inhabitants of Indiana spread across 23 health systems and 93 hospitals, as well as other state laboratory reporting state vitals data. The INPC patient population represents a variety of health systems spread across Indiana [10] (representation of COVID-19 patient dataset is illustrated in detail in the Results section). This is relevant given that Indiana is representative of the total US population in terms of age, gender, education levels [11] and urban-rural divide [12]. For each patient, CoRDaCo includes data captured between January 1, 2018, and November 30, 2020. The data pull was performed by specialized analysts from the Regenstrief Institute Data Core—the only personnel permitted direct access to identifiable patient data within the INPC research database.

Preparation of Feature Sets

We extracted and vectorized a wide variety of patient-level features representing their demographics; diagnoses; past encounter history; medications; and social determinants of health, defined as conditions in which people are born, grow, live, work and age [13] (Table 1).

Creation of feature vectors for model development was performed by the authors using the python programming language.

Table 1. List of features extracted for model development.

Data type	Description of features modeled
Demographics	Patient age, gender, race or ethnicity represented as integer and categorical variables
Diagnosis data	Represented as integer variables: • Charlson comorbidity index [14] Represented as a Boolean values: • Presence of most commonly occurring chronic conditions [15] • Diagnoses of addictions, behaviors, behavioral disorders, and narcotics use [16] • Presence of 1000 most frequently reported diagnoses identified using the International Classification of Diseases
Past encounter history	Inpatient, outpatient, and emergency visits represented as counts
Medications	Medications categorized into diagnosis groups and represented as Boolean values
Social determinants of health	Represented as a Boolean values: • Socioeconomic status (unemployment, type of insurance) • Education • Neighborhood and physical environment • Urban vs rural status classified using Rural-Urban commuting area (RUCA) codes [17] • Employment • Social support networks • Access to health care according to the Kaiser Family Foundation framework [18] All features were inferred using patient-level diagnosis codes and patient address information.

Development of a Gold Standard

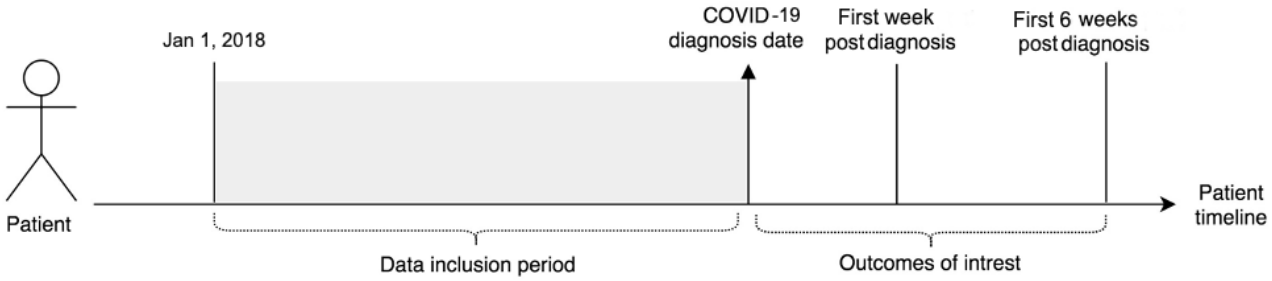
We parsed past encounter history data on each patient to identify those who had been hospitalized (defined as patients who had been admitted to either inpatient or intensive care) within either of the following:

- The first week of receiving a diagnosis of COVID-19 (ie, 1-week cohort), including a measure of which patients were in need of urgent care at the time of, or soon after, diagnosis.
- The first 6 weeks of receiving a diagnosis of COVID-19 (ie, 6-week cohort). A metric of which patients would need inpatient care during the course of their illness [19].

To ensure that our gold standard focused on inpatient or intensive care unit stays influenced by COVID-19 alone, we applied regular expressions to patient admission reason notes in order to identify and exclude any admissions due to accidents such as falls, injuries, lacerations, and fractures, as well as suicidal ideation, overdoses, and alcohol abuse. These factors were selected for exclusion based on an assessment of the most frequently occurring admission reasons identified from patient hospitalization datasets.

Figure 1 represents our approach to feature vector preparation and detection of outcomes of interest for analytical modelling based on the patient's longitudinal health history.

Figure 1. Feature vector preparation and detection of outcomes of interest based on the patient's longitudinal health history.



Machine Learning Process

We leveraged Python and the scikit-learn machine learning library [20] to train prediction models using the eXtreme Gradient Boosting (XGBoost) algorithm [21]. The XGBoost algorithm is an implementation of gradient-boosted ensemble decision trees [22] designed to optimize speed and performance. XGBoost classification was selected because research conducted by ourselves, as well as other external groups found that ensemble decision trees performed compatibly, or better than other classification algorithms [23,24] and because XGBoost could be trained using a smaller number of features than those required to train neural networks and other deep learning–based models, which enables ease of model development, interpretability, and explainability. We split each data vector into random groups of 80% (training and validation dataset) and 20% (holdout test set). We then leveraged the 80% training and validation dataset to train optimal models for each scenario by using 10-fold crossvalidation and hyperparameter turning and methods. To enable better generalization of each model, we applied the internal feature selection method of XGBoost [25], which prioritizes feature importance based on average

gain across all splits the feature is used in, to restrict models to a smaller subset of the most relevant features.

Model Evaluation

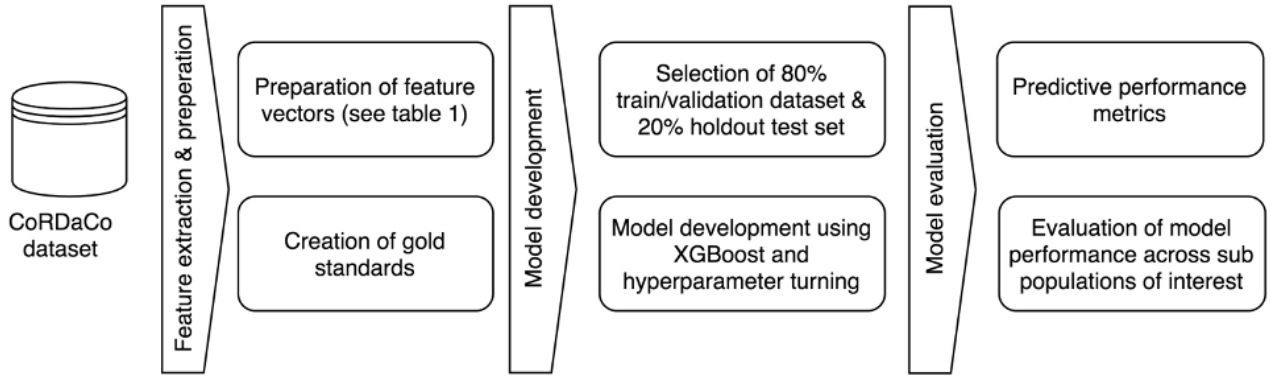
We assessed the performance of each decision model in the 20% holdout test dataset by using several performance metrics:

- Positive predictive value, or *precision*: the likelihood that a positively identified case is truly positive.
- Sensitivity, or *recall*: the likelihood that a true positive case is correctly identified as positive.
- Specificity: the likelihood that a negative case is correctly identified as negative.
- F_1 score: the harmonic mean of model precision and recall scores.
- Accuracy: the likelihood that a prediction is correct.
- Area under the receiver operating curve (AUC-ROC): a metric representing the performance of a prediction model at all classification thresholds.

Evaluation of Analytical Performance Against Subpopulations

As discussed previously, the COVID-19 pandemic has highlighted systemic disparities in patient outcomes and access to care based on race or ethnicity [4], gender [5], income level, and urban-rural divide [6,7]. These disparities may be present in the datasets used to train analytical models, resulting in biased predictions that place privileged groups at a systematic advantage and unprivileged groups at a systematic disadvantage [26]. To evaluate our models for such biases, we stratified the holdout test dataset by age, race or ethnicity, gender, and residence (urban vs rural), and we evaluated model performance across each stratified subpopulation by using the same performance metrics. Figure 2 provides a comprehensive overview of our study approach.

Figure 2. Workflow presenting the complete study approach from data extraction to predictive model evaluation. CoRDaCo: COVID-19 Research Data Commons.



Human Participants Research Approval

This study was approved by the Indiana University's Institutional Review Board (2005573466).

Results

Overview

The CoRDaCo dataset consisted of 230,981 patients with a positive COVID-19 diagnosis. However, we noted that a considerable number of these patients were out-of-state residents who visited health systems that were part of the INPC only to obtain COVID-19 tests or were Indiana residents whose only interaction with INPC-affiliated health systems were to undergo COVID-19 testing. As such, we had no clinical data beyond COVID-19 status on these patients. To enrich the quality of datasets used for model building, we excluded such patients by identifying and removing any patient whose only INPC record was a positive COVID-19 test result. This resulted in a total of 96,115 patients. We excluded an additional 89 patients owing to errors in their medical records, resulting in a total of 96,026 *legacy patients* to be included in our model development efforts. This legacy population was from a diverse race or ethnicity (27% Black, Hispanic, and others), predominantly adult (median age 47 years [33.73]), mostly urban (76,988/96,026, 80.17%), and had a larger representation of females (57,475/96,026, 59.85%). A total of 18,694 (19.47%) of these patients were hospitalized during the first week of being diagnosed with COVID-19, whereas 22,678 (23.62%) were hospitalized during the first 6 weeks of receiving a COVID-19 diagnosis.

Table 2. Characteristics of the patient populations used for analytical model development.

Patient characteristics	COVID-19 patient cohort	Patients hospitalized during the first week	Patients hospitalized during the first 6 weeks
Gender, n (%)			
Male	38,529 (40.12)	8178 (43.75)	9615 (42.40)
Female	57,475 (59.85)	10,516 (56.25)	13,062 (57.60)
Unknown	22 (0.02)	0 (0)	1 (0)
Race or ethnicity, n (%)			
White, non-Hispanic	70,238 (73.15)	11,783 (63.03)	14,737 (64.98)
Black, non-Hispanic	12,372 (12.88)	4,104 (21.95)	4666 (20.58)
Hispanic	9882 (10.29)	2171 (11.61)	2,533 (11.17)
Other	3534 (3.68)	636 (3.40)	742 (3.27)
Age (years), n (%)[a]			
Minors (<18 years)	7064 (7.36)	638 (3.41)	754 (3.34)
Adults (18-65 years)	67,563 (70.36)	11,330 (60.61)	13,851 (61.08)
Older adults (>65 years)	21,177 (22.05)	6726 (35.98)	8074 (35.60)
Unknown	222 (0.23)	0 (0)	0 (0)
Residence, n (%)			
Number of zip codes represented	957 (99.90)	678 (70.85)	705 (73.67)
Living in an urban area	76,988 (80.17)	14,833 (79.35)	17,910 (78.98)
Living in a rural area	16,843 (17.54)	3267 (17.48)	4084 (18.01)
Unknown	2195 (2.29)	594 (3.18)	684 (3.02)
Encounters, mean (SD)			
Outpatient visits	7.715 (10.09)	9.391 (13.18)	9.530 (12.29)
Emergency room visits	0.926 (2.25)	2.431 (3.52)	2.237 (3.45)
Hospitalizations	0.339 (1.35)	0.938 (2.24)	0.875 (2.19)
Chronic disease burden, n (%)			
Cancer	3976 (4.14)	1226 (6.56)	1484 (6.54)
Diabetes with complications	4340 (4.52)	1903 (10.18)	2222 (9.80)
Diabetes without complications	10,819 (11.27)	3845 (20.57)	4506 (19.87)
Dementia	2529 (2.63)	648 (3.47)	871 (3.84)
Chronic pulmonary disease	10,755 (11.20)	2364 (12.65)	4338 (19.13)
Renal disease	5449 (5.67)	2397 (12.82)	2794 (12.32)

[a]Mean participant age: 47.039 years (21.43).

Model Development and Evaluation

The feature preparation process (Table 1) resulted in a total of 1172 features for model training. To enable model generalizability and ease of interpretation, we restricted each model to approximately the most significant 100 features selected based on feature importance threshold drop-offs. Table 3 presents performance metrics reported by each model across the 20% holdout test dataset. Figure 3 presents the precision-recall and AUC-ROC curves for each prediction model. The subset of features included in each model is presented in Multimedia Appendix 1.

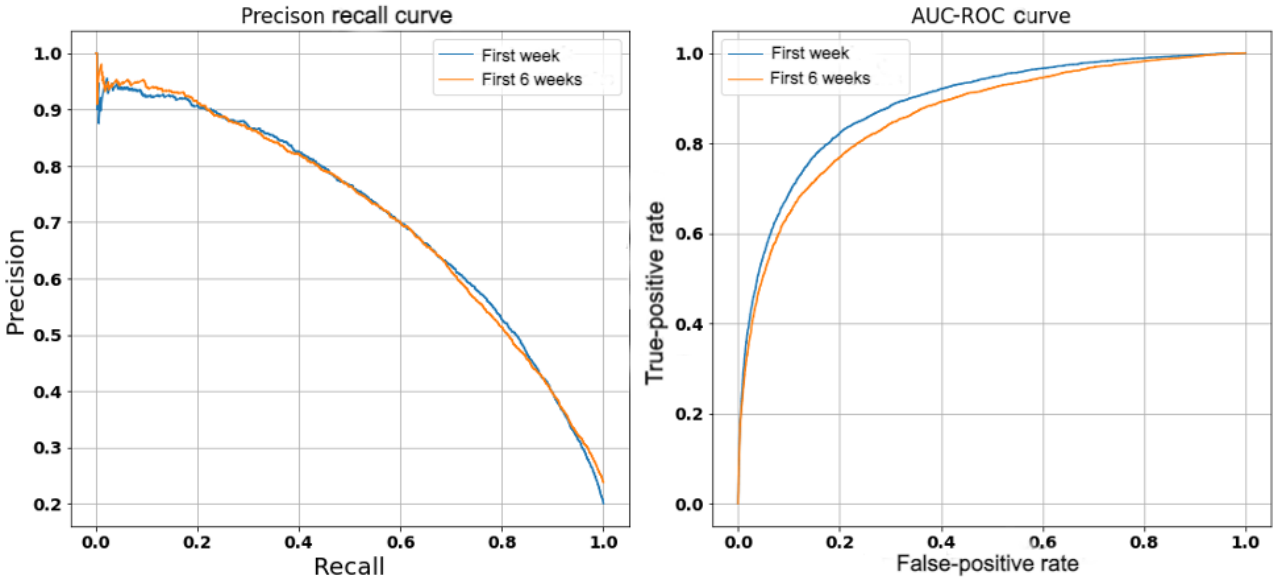
Both models delivered strong performance metrics. However, the model for the 1-week cohort reported significantly greater specificity, accuracy, and AUC-ROC scores than the 6-week cohort model.

Table 3. Predictive model performance.

Performance metric	First week (95% CI)	First 6 weeks (95% CI)
Precision	75.133 (73.445-76.822)	73.697 (72.142-75.253)
Sensitivity	52.505 (50.875-54.136)	52.571 (51.081-54.061)
Specificity	95.780 (95.457-96.104)	94.269 (93.887-94.653)
Accuracy	87.326 (86.846-87.806)	84.514 (83.992-85.037)
AUC-ROC[a]	88.744 (88.136-89.205)	86.215 (85.773-87.091)
F_1 score	61.814 (60.092-63.535)	61.367 (59.797-62.936)

[a]AUC-ROC: area under the receiver operating curve.

Figure 3. Precision-recall and AUC-ROC: area under the receiver operating curve (AUC-ROC) curves for each prediction model.



Evaluation of Analytical Performance Against Subpopulations

To assess model performance across different subpopulations of interest, we stratified the holdout test dataset by age, race or ethnicity, gender, and residence (urban vs rural), and we then evaluated their performance using each performance metric. Tables 4 and 5 present statistically significant variations in predictive performance reported across each model. Comprehensive predictive performance metrics, together with 95% CIs are listed in Multimedia Appendix 2. AUC-ROC curves for the performance of models across each stratified subpopulation are presented in Multimedia Appendix 3.

Table 4. Statistically significant performance variations in model to predict health care resource utilization within the first week.

Performance metric	Urban vs rural	Male vs female	Minors vs adults vs older adults	White vs Black vs Hispanic
Precision	No difference	No difference	No difference	No difference
Sensitivity or recall	Urban > rural	Male > female	Older adults > (adults = minors)	Black > Hispanic > White
Specificity	No difference	No difference	Minors > adults > older adults	(White and Hispanic) > black
Accuracy	No difference	No difference	Minors > adults > older adults	(White and Hispanic) > black
AUC-ROC[a]	No difference	Male > female	Minors > adults > older adults	No difference
F_1 score	Urban > rural	Male > female	(Older adults = minors) > adults	(Black and Hispanic) > White

[a]AUC-ROC: area under the receiver operating curve.

Table 5. Statistically significant performance variations in model to predict health care resource utilization within the first 6 weeks.

Performance metric	Urban vs rural	Male vs female	Minors vs adults vs older adults	White vs Black vs Hispanic
Precision	No difference	No difference	No difference	No difference
Sensitivity or recall	Urban > rural	Male > female	Older adults > (adults = minors)	Black > Hispanic > White
Specificity	No difference	No difference	Minors > adult > senior	White & Hispanic > black
Accuracy	No difference	No difference	Minors > adult > senior	No difference
AUC-ROC[a]	Urban > rural	Male > female	Minors > adult > senior	No difference
F_1 score	Urban > rural	Male > female	(Older adults = minors) > adults	Black > Hispanic > White

[a]AUC-ROC: area under the receiver operating curve.

As presented in Tables 4 and 5, there were no statistically significant differences in precision scores reported across each strata or model under test. However, we found evidence of significant variations in model performance across many other strata. Across both models and all performance metrics under test, residing in an urban area was associated with comparable, or higher predictive performance than if residing in a rural area. Across both models and all performance metrics under test, being male was associated with comparable, or higher predictive performance than if female. Performance stratified by age showed significant variations, with some performance metrics favoring older adults while others favored minors. These results are indicative of biases learned from underlying data sources used for model development, or inefficient learning parameters implemented by the machine learning algorithm.

Discussion

Principal Findings

Our results demonstrate the ability to train decision models capable of predicting the need of COVID-19–related hospitalization across a broad, statewide patient population with considerable performance accuracy. The 1-week model for predicting the need of COVID-19–related hospitalization reported specificity, accuracy, and AUC-ROC scores that were significantly larger than the 6-week model. The findings are intuitive given that hospitalization risk is more predictable over shorter time frames. Such utilization prediction models may be used for population health management programs in health systems, to identify high-risk populations to monitor or screen, as well as predicting resource need in crisis situations, such as future spikes in pandemic activity or outbreaks.

Stratification of model performance across age, race or ethnicity, gender, and urban versus rural divide identified statistically significant variations in model performance across subpopulations. Each model and stratified subpopulation under test reported precision scores >70%, accuracy and AUC-ROC scores >80%, and sensitivity scores approximately >90%. We note that recall scores for each model (approximately 50%-54%) were lower than ideal, implying that a considerable proportion of patients in need of health care services were being ignored. However, model precision, which is indicative of what percentage of patients identified by the model actually needed care was high (>70%), suggesting that it was pragmatic for use in clinical settings. Additionally, model specificity scores were very high (approximately >90%). This finding indicated that the models were able to correctly identify patients who were not in need of care with very high accuracy, which is very valuable in making clinical decisions on which patients to prioritize.

Features that influenced the prediction of health care resource utilization included patient age [27], chronic obstructive pulmonary disease status [28], smoking [28], diabetes [29], indication of neurological diseases via diagnosis (eg, dementia [30]) or medications (eg, anti-Parkinson and related therapy agents), mental disorders (eg, anxiety disorders), residence (urban vs rural) [31,32], and income-level, measured on the basis of the type of insurance used by the patient. None of the patient-level social determinants of health factors extracted from the International Classification of Diseases diagnosis data were found to be impactful enough for inclusion in either model. This could be attributed to the scarcity of these elements being captured in clinical settings. However, patient-level features on the type of insurance (which is indicative of an individual's financial and employment status) and RUCA code (which could be used to infer an individual's income level, isolation, and access to services and health resources) were both widely available. These elements were found to be impactful and were integrated into both models.

Each model exhibited significant variations in predictive performance across subpopulations. Overall, male gender or living in an urban area was associated with stronger predictive performance. These differences may be influenced by variations in access to health care services or health care delivery prevalent in the datasets, and the models could learn them during the training process. We cannot make further assumptions on the causes of varying model predictions without a proper assessment of underlying causes of this behavior.

Limitations

We noted several limitations to this study. We leveraged statewide datasets from the INPC HIE system to ensure that our models could be operationalized across a broad geographic region. As such, our modeling did not include data elements that were collected by health systems but not shared with the INPC. Since the collection of such datasets and their availability at the HIE level may vary based on the health system, the inclusion of such elements may impact the generalizability of our models across different health systems. Our use cases assessed the need of hospitalization during the first 6 weeks of

diagnosis. This excludes the needs of patients suffering long-COVID, where patients may not fully recover for several months [33]. Models were trained using *legacy patients*, who were participants of the INPC system prior to March 1, 2020. It is unclear how the models will perform against other patients who do not regularly interact with the health system and sought care only for COVID-19 testing purposes. This is concerning given that such patients may suffer from a higher disease burden. Our modelling efforts covered a broad time period spanning several waves of the COVID-19 pandemic, as well as the enforcement and relaxation of various mandates aimed at controlling COVID-19 infection rates. These changes may have influenced the capacity of hospital systems resulting in changes in how many patients were provided inpatient care. Alternatively, hospital admission and emergency management protocols may have also changed throughout this period, further impacting which patients received care. Our current effort did not consider how these variations influence the training datasets, and as such, how our models would generalize across future outbreaks and mandates, as COVID-19 infection rates continue to change. Future research will systematically investigate and calibrate model performance across different stages of the pandemic.

We sought to demonstrate the ability to develop broad, state-level models for COVID-19–related research. As such, the biases in analytical models detected in this study highlight significant concerns that researchers must protect against. These biases in analytical model performance will be addressed during the next phase of our work. Further, although the generalizability of our models across other states is untested, they can influence other emerging COVID-19 analytical efforts. In particular, these models can influence data collection, curation, and modeling activities undertaken by the National COVID Cohort Collaborative (N3C) [34], which is stewarded by the National Center for Advancing Translational Sciences and hosts data on over 250,000 COVID-19–positive patients from 31 sites spread across the United States. N3C could serve as an in-vivo laboratory for our research efforts.

Conclusions

This study presents the possibility of developing decision models capable of predicting patient-level health care resource utilization across a broad, statewide region with considerable predictive performance. However, the analytical models present statistically significant variations in performance across stratified subpopulations of interest. Further efforts are necessary to identify root causes of these biases and to rectify them.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of top-ranking features included in each predictive model.
[DOCX File , 15 KB - jmir_v23i11e31337_app1.docx]

Multimedia Appendix 2
Performance metrics reported by each analytical model across each stratified subpopulation of the study.
[DOCX File , 18 KB - jmir_v23i11e31337_app2.docx]

Multimedia Appendix 3
Area under the receiver operating curves (AUC-ROCs) for the performance of models across each stratified subpopulation.
[PNG File , 261 KB - jmir_v23i11e31337_app3.png]

References

1. Estiri H, Strasser ZH, Klann JG, Naseri P, Wagholikar KB, Murphy SN. Predicting COVID-19 mortality with electronic medical records. NPJ Digit Med 2021 Mar 04;4(1):15 [FREE Full text] [doi: 10.1038/s41746-021-00383-x] [Medline: 33542473]
2. Hartley DM, Perencevich EN. Public Health Interventions for COVID-19: Emerging Evidence and Implications for an Evolving Public Health Crisis. JAMA 2020 May 19;323(19):1908-1909. [doi: 10.1001/jama.2020.5910] [Medline: 32275299]
3. Liu Q, Luo D, Haase JE, Guo Q, Wang XQ, Liu S, et al. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. The Lancet Global Health 2020 Jun;8(6):e790-e798. [doi: 10.1016/s2214-109x(20)30204-7]
4. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. The Lancet 2020 Apr;395(10232):1243-1244. [doi: 10.1016/s0140-6736(20)30893-x]
5. Fisher AN, Ryan MK. Gender inequalities during COVID-19. Group Processes & Intergroup Relations 2021 Mar 04;24(2):237-245. [doi: 10.1177/1368430220984248]
6. Iyanda AE, Boakye KA, Lu Y, Oppong JR. Racial/ethnic heterogeneity and rural-urban disparity of COVID-19 case fatality ratio in the USA: a negative binomial and GIS-based analysis. J Racial Ethn Health Disparities 2021 Mar 26:1-14 [FREE Full text] [doi: 10.1007/s40615-021-01006-7] [Medline: 33638102]

7. Huang Q, Jackson S, Derakhshan S, Lee L, Pham E, Jackson A, et al. Urban-rural differences in COVID-19 exposures and outcomes in the South: A preliminary analysis of South Carolina. PLoS One 2021 Feb 3;16(2):e0246548 [FREE Full text] [doi: 10.1371/journal.pone.0246548] [Medline: 33534870]
8. Regenstrief – IU partnership offers fast, secure access to COVID-19 data for research. Regenstrief Institute. 2021. URL: https://www.regenstrief.org/article/regenstrief-iu-cordaco-partnership-announcement/ [accessed 2021-09-30]
9. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, INPC Management Committee. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health Aff (Millwood) 2005 Sep;24(5):1214-1220. [doi: 10.1377/hlthaff.24.5.1214] [Medline: 16162565]
10. Indiana Health Information Exchange – Overview. 2017 Mar. URL: https://www.ihie.org/wp-content/uploads/2017/05/IHIE-Overview_March-2017.pdf [accessed 2021-09-30]
11. Quick Facts, United States. United States Census Bureau. 2021. URL: https://www.census.gov/quickfacts/fact/table/US/RHI525219 [accessed 2021-09-30]
12. State Fact Sheets. Economic Research Service - U.S. Department of Agriculture. URL: https://www.ers.usda.gov/data-products/state-fact-sheets/ [accessed 2021-09-30]
13. Social determinants of health. World Health Organization. URL: https://www.who.int/social_determinants/sdh_definition/en/ [accessed 2021-09-30]
14. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. J Clin Epidemiol 2008 Dec;61(12):1234-1240. [doi: 10.1016/j.jclinepi.2008.01.006] [Medline: 18619805]
15. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Prev Chronic Dis 2013 Apr 25;10:E66 [FREE Full text] [doi: 10.5888/pcd10.120239] [Medline: 23618546]
16. Dominion Diagnostics. URL: https://www.dominiondiagnostics.com/sites/default/files/ICD-10-CM%20Addiction%20Med%20%26%20Behavioral%20Health%202017.pdf [accessed 2021-09-30]
17. Rural-Urban Commuting Area Codes (2021). U.S. Department of Agriculture – Economic Research Service. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2021-09-30]
18. Artiga S, Hinton E. Beyond health care: the role of social determinants in promoting health and health equity 2018. Henry J Kaiser Family Foundation. 2018 May. URL: http://files.kff.org/attachment/issue-brief-beyond-health-care [accessed 2021-09-30]
19. 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 2021-09-30]
20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. Journal of Machine Learning Research 2011 Oct;12:2825-2830.
21. Chen T, Guestrin C, editors. XGBoost: a scalable tree boosting system. : Association for Computing Machinery; 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
22. Ye J, Chow J, Chen J, Zheng Z. Stochastic gradient boosted distributed decision trees. : Association for Computing Machinery; 2009 Presented at: Proceedings of the 18th ACM Conference on Information and Knowledge Management; November 2009; Hong Kong, China p. 2061-2064. [doi: 10.1145/1645953.1646301]
23. Kasthurirathne SN, Grannis S, Halverson PK, Morea J, Menachemi N, Vest JR. Precision health-enabled machine learning to identify need for wraparound social services using patient- and population-level data sets: algorithm development and validation. JMIR Med Inform 2020 Jul 09;8(7):e16129 [FREE Full text] [doi: 10.2196/16129] [Medline: 32479414]
24. Didrik N. Tree Boosting With XGBoost - Why Does XGBoost Win "Every" Machine Learning Competition? Norwegian University of Science and Technology. 2016. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2433761 [accessed 2021-09-30]
25. Python API Reference. XGBoosT. URL: https://xgboost.readthedocs.io/en/latest/python/python_api.html?highlight=get_score#xgboost.Booster.get_score [accessed 2021-09-30]
26. Ferryman K, Pitcan M. Fairness in precision medicine. Data & Society. 2018. URL: https://datasociety.net/library/fairness-in-precision-medicine/ [accessed 2021-09-30]
27. Lithander F, Neumann S, Tenison E, Lloyd K, Welsh T, Rodrigues J, et al. COVID-19 in older people: a rapid clinical review. Age Ageing 2020 Jul 01;49(4):501-515 [FREE Full text] [doi: 10.1093/ageing/afaa093] [Medline: 32377677]
28. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. J Med Virol 2020 Oct 17;92(10):1915-1921 [FREE Full text] [doi: 10.1002/jmv.25889] [Medline: 32293753]
29. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev 2020 Mar 31:e3319 [FREE Full text] [doi: 10.1002/dmrr.3319] [Medline: 32233013]

30. Zuin M, Guasti P, Roncon L, Cervellati C, Zuliani G. Dementia and the risk of death in elderly patients with COVID‐19 infection: Systematic review and meta‐analysis. Int J Geriatr Psychiatry 2020 Nov 24;36(5):697-703. [doi: 10.1002/gps.5468]
31. Souch JM, Cossman JS. A commentary on rural-urban disparities in COVID-19 testing rates per 100,000 and risk factors. J Rural Health 2021 Jan;37(1):188-190 [FREE Full text] [doi: 10.1111/jrh.12450] [Medline: 32282964]
32. Zhang J, Zhu L, Li S, Huang J, Ye Z, Wei Q. Rural-urban disparities in knowledge, behaviors, and mental health during COVID-19 pandemic: A community-based cross-sectional survey. Medicine 2021;100(13):e25207. [doi: 10.1097/md.0000000000025207]
33. Nabavi N. Long covid: How to define it and how to manage it. BMJ 2020 Sep 07;370:m3489. [doi: 10.1136/bmj.m3489] [Medline: 32895219]
34. National COVID Cohort Collaborative (N3C). National Institutes of Health. URL: https://ncats.nih.gov/n3c [accessed 2021-09-30]

Abbreviations

AUC-ROC: Area under the receiver operating curve
CoRDaCo: COVID-19 Research Data Commons
HIE: Health Information Exchange
INPC: Indiana Network for Patient Care
N3C: National COVID Cohort Collaborative
RUCA: Rural-Urban commuting area
XGBoost: eXtreme Gradient Boosting

Edited by C Basch; submitted 20.06.21; peer-reviewed by A Khurshid, S Nagavally, I Mircheva, Z Xie; comments to author 19.07.21; revised version received 07.09.21; accepted 15.09.21; published 15.11.21.

Please cite as:
Kasturi SN, Park J, Wild D, Khan B, Haggstrom DA, Grannis S
Predicting COVID-19–Related Health Care Resource Utilization Across a Statewide Patient Population: Model Development Study
J Med Internet Res 2021;23(11):e31337
URL: https://www.jmir.org/2021/11/e31337
doi:10.2196/31337
PMID:34581671

Original Paper

A Markerless 2D Video, Facial Feature Recognition–Based, Artificial Intelligence Model to Assist With Screening for Parkinson Disease: Development and Usability Study

Xinyao Hou[1*], BA; Yu Zhang[2*], MD, PhD; Yanping Wang[3*], MD, PhD; Xinyi Wang[4*], MA; Jiahao Zhao[2], MA; Xiaobo Zhu[2], BA; Jianbo Su[1], PhD

[1]Department of Automation, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Neurology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of Neurology, Second Affiliated Hospital of Jiaxing City, Jiaxing, China

[4]Department of Neurology, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jianbo Su, PhD
Department of Automation
Shanghai Jiao Tong University
800 Dongchuan Road, Minhang District
Shanghai, 200240
China
Phone: 86 21 34204276
Email: jbsu@sjtu.edu.cn

Abstract

Background: Masked face is a characteristic clinical manifestation of Parkinson disease (PD), but subjective evaluations from different clinicians often show low consistency owing to a lack of accurate detection technology. Hence, it is of great significance to develop methods to make monitoring easier and more accessible.

Objective: The study aimed to develop a markerless 2D video, facial feature recognition–based, artificial intelligence (AI) model to assess facial features of PD patients and investigate how AI could help neurologists improve the performance of early PD diagnosis.

Methods: We collected 140 videos of facial expressions from 70 PD patients and 70 matched controls from 3 hospitals using a single 2D video camera. We developed and tested an AI model that performs masked face recognition of PD patients based on the acquisition and evaluation of facial features including geometric and texture features. Random forest, support vector machines, and k-nearest neighbor were used to train the model. The diagnostic performance of the AI model was compared with that of 5 neurologists.

Results: The experimental results showed that our AI models can achieve feasible and effective facial feature recognition ability to assist with PD diagnosis. The accuracy of PD diagnosis can reach 83% using geometric features. And with the model trained by random forest, the accuracy of texture features is up to 86%. When these 2 features are combined, an F1 value of 88% can be reached, where the random forest algorithm is used. Further, the facial features of patients with PD were not associated with the motor and nonmotor symptoms of PD.

Conclusions: PD patients commonly exhibit masked facial features. Videos of a facial feature recognition–based AI model can provide a valuable tool to assist with PD diagnosis and the potential of realizing remote monitoring of the patient's condition, especially during the COVID-19 pandemic.

(J Med Internet Res 2021;23(11):e29554) doi:10.2196/29554

KEYWORDS

Parkinson disease; facial features; artificial intelligence; diagnosis

Introduction

Parkinson disease (PD) is a typical movement disorder, and its high rate of disability seriously affects the daily life of patients [1]. However, there is a lack of reliable methods for the early diagnosis of PD. The diagnostic accuracy of the clinical diagnosis of PD has been reported at 73.8% when performed mainly by nonexperts and 79.6% when performed by movement disorder experts [2]. Hence, it is important to develop a valuable tool to diagnose PD as early as possible. In recent years, computer technology and artificial intelligence (AI) have been making great advances in the diagnosis of PD. Many studies have been carried out by using computer technology from different aspects [3], including using wearable devices to study the gait of patients with PD [4] and using intelligent pens to study the trembling of patients' writing based on their hand tremor [5]. However, the accuracy of these studies is not sufficient, and they cannot be applied on a large scale due to the limitations of equipment and technology.

Patients with PD usually suffer from loss of facial expressions. Their facial muscle movements and amplitude of expression are usually different from healthy people because of facial muscle stiffness, which is called masked face [6]. Based on this, research on the facial expressions of PD when watching videos and during social interactions has been reported [7]. The overall expressivity of PD has been verified as being reduced, but subjective evaluations from different clinicians often show low consistency [8]. And traditional methods are not accessible and fast enough for large-scale work. Computer vision has been considered to assist the work.

The use of 2D video to assist with medical diagnosis has long-standing precedents. In 2014, researchers used independent component analysis and a number of different classifiers to describe local shape variations and differentiate people with Down syndrome from the general population [9]. Basel-Vanagaite et al [10] used novel facial dysmorphology analysis to identify patients with Cornelia de Lange syndrome and achieved a good effect, with a recognition rate of 87%. Studies on facial expression recognition algorithms have also emerged in psychiatry [11]. These applications all show that 2D video is effective for diagnosis when the disease has an impact on facial expressions. The use of computer vision on prescreening can not only quickly process a large amount of data and reduce the burden on doctors but also provide reliable support to patients remotely. Now that the COVID-19 pandemic is ravaging, guidelines for avoiding contact can undoubtedly better support epidemic prevention and control. Therefore, we aimed to explore a convenient and accessible PD diagnostic method using a markerless 2D video, facial feature recognition–based, AI model.

In our study, the facial features of PD patients and nonpatients were collected that represent the speed, elasticity, and coordination of the facial muscles [12]. Our goal was to identify features from facial information that distinguish PD from healthy people and finally develop a markerless 2D video, facial feature recognition–based, AI model to assist with PD diagnosis.

Methods

Study Design and Participants

We performed a multicenter, observational, outpatient-based, cross-sectional study, which was approved by the Research Ethics Committee of Xin Hua Hospital affiliated with Shanghai Jiao Tong University School of Medicine and the Research Ethics Committee of each site of the study group. After obtaining informed consent, participants were consecutively recruited. PD was diagnosed according to the Movement Disorder Society PD criteria. Patients with secondary parkinsonism, stroke, brain tumor, or an alternative cause for parkinsonism symptoms were excluded. For the control group, age- and gender-matched healthy participants were recruited. The sample size of our study was determined using a priori statistical power analysis. On the basis of the literature, we predicted a large between-subjects effect size in our design using analysis of variance (ANOVA; F=0.4). A power analysis using G-Power indicated that a total sample size of 44 (22 per group) would be needed, with 95% power (1-β), an alpha of .05, and a correlation of 0.5 among the 120 repeated measurements. These analyses suggest that our recruited sample of 140 participants (70 patients, 70 controls) would be sufficient.

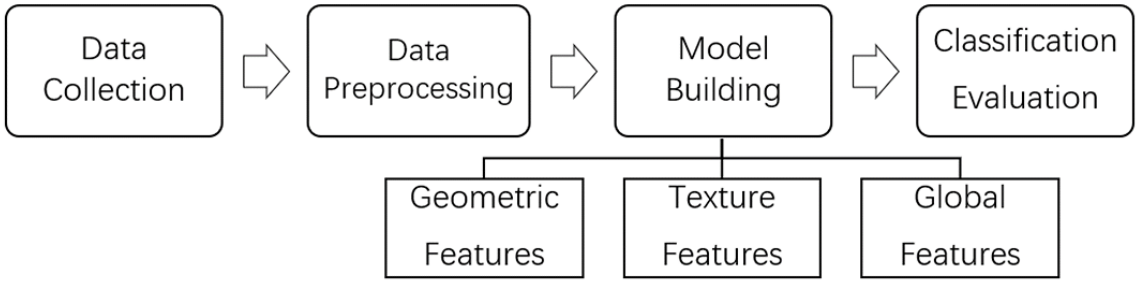
Data Collection

A total of 70 consecutive PD patients (male/female: 33/37) and 70 matched controls (MC; male/female: 39/31) were included in our study. In this study, a structured interview for clinical and demographic variables was performed. All PD participants were evaluated during the medication "on" period. The following demographic and clinical feature data were obtained from the participants: gender, age, education, age at PD onset, duration of disease, the history of dopamine replacement therapy and related complications (total levodopa equivalent daily dose [LEDD] was calculated according to previously suggested conversion formulae), Hoehn-Yahr (H-Y) stage, Unified Parkinson Disease Rating Scale (UPDRS), Mini-Mental State Examination (MMSE), a scale for freezing of gait, Non-Motor Symptoms Scale (NMSS), REM Sleep Behavior Disorder Questionnaire Hong Kong (RBDQ-HK), Hamilton Anxiety Scale (HAM-A), Parkinson Disease Questionaire-39 (PDQ-39), and the Questionnaire for Impulsive-Compulsive Disorders in Parkinson disease (QUIP).

The flowchart of the research process is shown in Figure 1 and included the following steps: Collect a series of facial expressions from patients and nonpatients, and perform a series of preprocessing operations to improve the quality of the samples. The corresponding models were constructed by extracting geometric features and texture features. Finally, these 2 features were combined as combined features to establish a patient identification model of PD, and the classification results of the model were evaluated.

The collected videos lasted 10-15 seconds each, and the patient involved was instructed to make certain expressions, including poker face and smiling, and was recorded by the camera. For each participant, we collected one such corresponding video.

Figure 1. Flowchart of the research process.



Preprocessing of Facial Images

The main work of preprocessing was the extraction of facial areas and localization and normalization of facial key points. Since information for facial expressions would be used, the ideal output of the preprocessing was the pure expressions. The preprocessing of facial images for this paper included the following 3 steps: (1) identify the existing face in the sample picture, and select the recognized face with a rectangular box; (2) locate the facial feature points in the intercepted rectangular box using 68 feature points, based on ensemble of regression trees cascade regression [13], to describe the position relationship of each key part of the face in detail; (3) use an affine transformation matrix to normalize the face to achieve the goal of face correction, which was obtained by the corresponding coordinate relationship between the feature points. In our study, face recognition and facial landmark detection were realized based on Dlib, a machine learning, open-source library. The preprocessing of sample images was processed through this series of processes to obtain the normalized facial image.

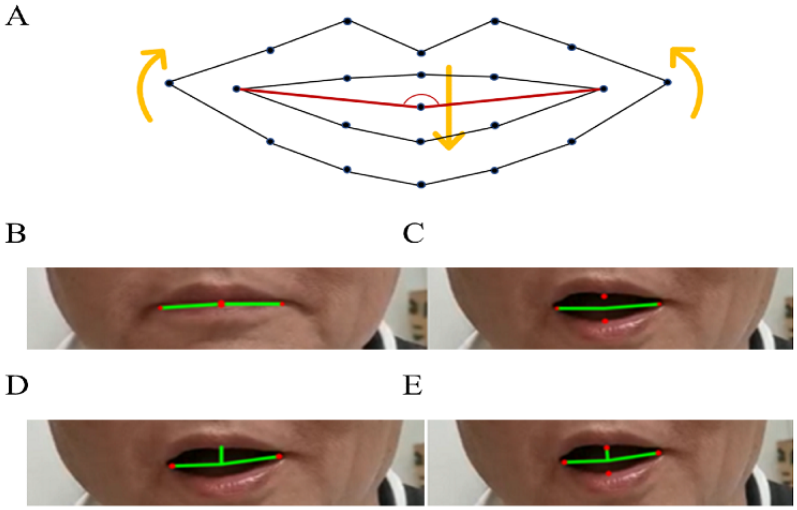
Extracting Facial Geometric Features

In this section, features were extracted from specific frames of the collected video, during which patients were poker-faced or smiling. The difference between these 2 frames was considered as features for further use. Due to the presence of “masked face” features in PD and the obvious appearance in the area of the mouth, the characteristic angle of the mouth was extracted using the key points of the mouth. Figure 2A shows the key points of the mouth and their movement direction during a smile, as well as the marking number of each key point in the detection process. The red lines construct the characteristic triangle of the mouth in this paper.

We defined 3 feature angles according to these feature points: (1) the angles before and after laughter, as shown in Figure 2B and Figure 2C; (2) the overall deviation angle of the mouth, as shown in Figure 2D; (3) the deviation angle of the left and right sides of the mouth, as shown in Figure 2E.

The feature angles constructed in our study can represent the overall range of motion of the mouth, so that the difference between the feature angle before and after a smile can reflect the stiffness of the mouth muscles in patients with PD, which can be used as a main feature. At the same time, the overall deviation angle of the mouth and the deviation angles of the left and right sides of the mouth can fully reflect the mouth asymmetry caused by uncoordinated mouth muscle movements in some patients with PD, so these geometric features can be used as an auxiliary feature. Thus, these features were mixed to construct a model to differentiate patients with PD by using geometric features.

Figure 2. The characteristic angle of the mouth. (A) movement and the relative position of key points on the mouth. (B) features of the angle before smiling. (C) features of the angle after smiling. (D) deviation angle of the overall mouth after smiling. (E) deviation angle of the left and right sides of the mouth after smiling.



Extracting Facial Texture Features

Compared with geometric features, texture features retain more important details because of the multiple dimensions. Moreover, facial texture is a relatively intuitive existence. For different people, texture itself has distinct characteristics, which can be expressed through features. In our study, the histogram of oriented gradient (HOG) [14,15] and local binary pattern (LBP) features [16,17] were used to describe facial textures. First, all samples were preprocessed. In order to further reduce the influence of irrelevant factors, the mouth and eye regions were selected as the feature regions and extracted for feature extraction. HOG features and LBP features were extracted from the obtained feature region [18].

For classification of texture features, this study used support vector machine (SVM) [19,20], k-nearest neighbor (KNN) [21,22], and random forest (RF; Tree) [23,24] to establish a recognition model to distinguish patients from people without PD. And we compared the results obtained by the different classifier algorithms to verify that the texture features of PD patients' faces are indeed different from those of MCs [25].

Statistical Analysis

All characteristics of our study were summarized using means and SDs for continuous variables and percentages and frequencies for categorical variables. The comparisons were performed using Student *t* tests to analyze continuous variables with parametric distributions and Mann-Whitney U tests to test variables with nonparametric distributions. And the Pearson chi-square test was used for categorical variables. Logistic regression analyses were adjusted for gender, age, age at PD onset, disease duration, and educational level. The level of significance was set at $P<.05$. Odds ratios (ORs) are presented with their 95% CIs. We used receiver operating characteristic (ROC) [26] analysis to assess the effectiveness of the classification. For the statistical analysis and to generate graphs, SPSS v24 (IBM Corp, Armonk, NY) and Prism 8.0 for Windows (GraphPad Software Inc, San Diego, CA) were used.

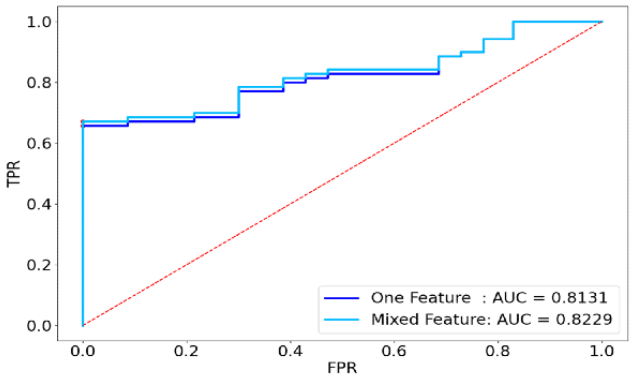
Results

Geometric Features of PD

For the geometric features, the samples were divided into a training set containing 80 sets of data and a testing set containing 60 sets of data. The area under the curve (AUC) of the model was 0.8131 for the main feature and 0.8229 for the mixed feature; values above 80% mean it has good performance for this problem (Figure 3A). The best threshold, Youden index from the ROC curve, and relevant F-measure and other indicators are shown in Figure 3B. The overall recognition effect of geometric features was about 80%, and the precision was 100%, while the recall was only 67%. However, the geometric feature model was not associated with clinical characteristics of PD (Table 1).

Figure 3. Receiver operating characteristics (ROC) analysis for the diagnosis of Parkinson disease using geometric feature and the recognition result. (A) ROC curve for each parameter and (B) result of machine learning algorithms. AUC: area under the curve; FPR: false positive rate; TPR: true positive rate.

A



B

Method	Precision	Recall	F-Measure	Accuracy
Geometric	1.00	0.67	0.80	0.83

Table 1. Correlations between clinical characteristics of Parkinson disease (PD) and geometric features of the face.

Variable	*P* value	r	95% CI
Age	.28	–0.131	–0.355 to 0.108
Age at onset	.08	–0.212	–0.426 to 0.024
PD duration	.12	0.19	–0.048 to 0.406
DALEDD[a]	.12	0.188	–0.049 to 0.405
QUIP[b]	.35	–0.114	–0.34 to 0.124
HAM-A[c]	.38	0.106	–0.132 to 0.333
RBD[d]	.53	–0.076	–0.305 to 0.0162
Freezing of gait	.73	0.042	–0.195 to 0.274
Total UPDRS[e]	.31	0.123	–0.115 to 0.348
HY[f]	.46	0.089	–0.149 to 0.318
MMSE[g]	.15	0.172	–0.065 to 0.391
NMSS[h]	.06	–0.223	–0.436 to 0.012
PDQ-39[i]	.54	0.074	–0.164 to 0.303

[a]DALEDD: levodopa equivalent daily doses.

[b]QUIP: Questionnaire for Impulsive-Compulsive Disorders in Parkinson disease.

[c]HAM-A: Hamilton Anxiety Scale.

[d]RBD: REM Sleep Behavior Disorder.

[e]UPDRS: Unified Parkinson Disease Rating Scale.

[f]HY: Hoehn & Yahr.

[g]MMSE: Mini-Mental State Examination.

[h]NMSS: Non-Motor Symptoms Scale.

[i]PDQ-39: Parkinson Disease Questionaire-39.

Texture Features of PD

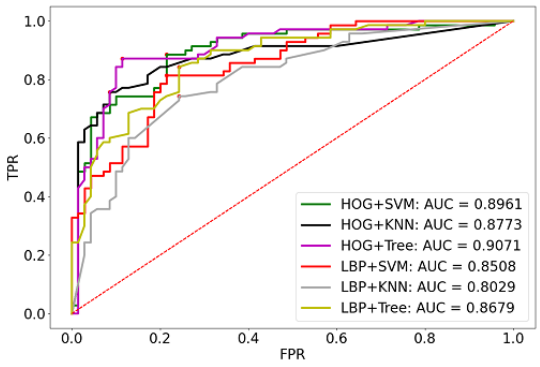
For the texture features, the AUC for LBP+KNN achieved 0.8029. The HOG+SVM (AUC=0.8961) and HOG+Tree (AUC=0.9071) methods performed the best. Regarding the texture feature extraction algorithms, the comprehensive performance of HOG was better than that of LBP, as the AUCs of HOG were all greater than 0.8773 (Figure 4A). The best threshold, Youden index from ROC curve, and relevant F-measure and other indicators are shown in Figure 4B. Because of the similar AUC results of HOG+SVM and HOG+Tree, we report the results from both. The overall recognition effect of HOG+Tree could reach 0.86, which means this method has excellent performance for this identification problem. While the recall of HOG+SVM was 0.87, most of the PD patients could be recognized under this method. And the F-measure of the methods was approximately 85%.

Since HOG+SVM and HOG+Tree performed well and had similar results, we further examined whether these texture features were significantly correlated with the clinically evaluated variables from the PD patients. It can be seen from Table 2 that the results of the texture features model were not associated with motor and nonmotor symptoms in PD.

To further explore the differences between the features of the eye and mouth in PD, we preprocessed the eye and mouth regions and obtained the corresponding texture features. Then, we compared the results of the diagnosis using the HOG+Tree method to explore the differences between them. It can be seen from Figure 4C that the AUC of the eye was 0.8955, which is higher than the 0.8799 of the mouth. And the values for the eye are better than those for the mouth for other evaluation indices (Figure 4D).

Figure 4. Receiver operating characteristics (ROC) analysis for the diagnosis of Parkinson disease using texture features and the recognition result. (A) ROC curve for 2 texture feature extraction algorithms with 3 classification models and (B) best result of machine learning algorithms. (C) ROC curve for texture features of the region of the eye, region of the mouth and their combination and (D) result of machine learning algorithms. AUC: area under the curve; FPR: false positive rate; HOG: histogram of oriented gradient; KNN: k-nearest neighbor; LBP: local binary patern; SVM: support vector machine; TPR: true positive rate.

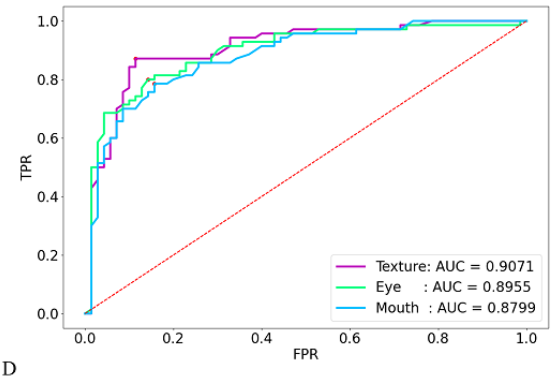
A



B

Method	Precision	Recall	F-Measure	Accuracy
HOG+SVM	0.80	0.87	0.84	0.83
HOG+Tree	0.88	0.84	0.86	0.86

C



D

Method	Precision	Recall	F-Measure	Accuracy
Texture	0.88	0.84	0.86	0.86
Eye	0.86	0.77	0.82	0.82
Mouth	0.83	0.77	0.80	0.81

Table 2. Correlations between the clinical characteristics and 2 main texture features.

Variable	HOG[a]+SVM[b]			HOG+Tree[c]		
	P value	r	95% CI	*P* value	r	95% CI
Age	.21	–0.152	–0.373 to 0.086	.13	–0.184	–0.402 to 0.053
Age at onset	.15	–0.175	–0.394 to 0.063	.26	–0.136	–0.359 to 0.102
PD[d] duration	.72	–0.043	–0.275 to 0.194	.15	–0.175	–0.394 to 0.062
DALEDD[e]	.75	0.039	–0.198 to 0.272	.65	–0.055	–0.286 to 0.183
QUIP[f]	.22	0.15	–0.088 to 0.372	.18	0.161	–0.077 to 0.382
HAM-A[g]	.38	0.106	–0.132 to 0.333	.96	–0.005	–0.24 to 0.23
RBD[h]	.27	0.134	–0.104 to 0.358	.22	0.15	–0.088 to 0.372
Freezing gait	.90	–0.016	–0.25 to 0.22	.94	–0.009	–0.244 to 0.226
Total UPDRS[i]	.75	0.039	–0.197 to 0.272	.58	–0.067	–0.298 to 0.17
HY[j]	.87	–0.02	–0.254 to 0.216	.74	–0.04	–0.273 to 0.196
MMSE[k]	.60	–0.064	–0.295 to 0.173	.97	–0.004	–0.239 to 0.231
NMSS[l]	.09	0.203	–0.034 to 0.418	.38	0.106	–0.132 to 0.333
PDQ39[m]	.95	–0.008	–0.242 to 0.228	.50	–0.083	–0.312 to 0.155

[a]HOG: histogram of oriented gradients.

[b]SVM: support vector machine.

[c]Tree: random forest.

[d]PD: Parkinson disease.

[e]DALEDD: levodopa equivalent daily dose.

[f]QUIP: Questionnaire for Impulsive-Compulsive Disorders in Parkinson disease.

[g]HAM-A: Hamilton Anxiety Scale.

[h]RBD: REM Sleep Behavior Disorder.

[i]UPDRS: Unified Parkinson Disease Rating Scale.

[j]HY: Hoehn & Yahr.

[k]MMSE: Mini-Mental State Examination.

[l]NMSS: Non-Motor Symptoms Scale.

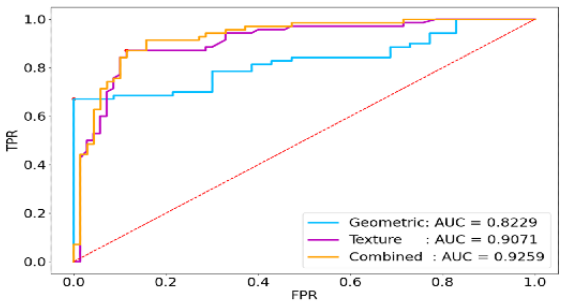
[m]PDQ-39: Parkinson Disease Questionnaire-39.

Combined Features of PD

For the combined features, it was higher (AUC 0.9259) than for geometric features alone (AUC 0.8229) and for texture features alone (AUC 0.9071), shown in Figure 5A. Similar to previous work, we obtained a Youden index as the corresponding optimal threshold from the ROC curve. Figure 5B shows that the final recognition effect of the combined features could reach 0.88, nearly 90%, with each data point as well as the comprehensive performance being improved to a certain extent.

Figure 5. Comparison of receiver operating characteristics (ROC) analysis for the diagnosis of Parkinson disease. (A) ROC curve for each parameter and (B) result of machine learning algorithms. AUC: area under the curve; FPR: false positive rate; TPR: true positive rate.

A



B

Method	Precision	Recall	F-Measure	Accuracy
Geometric	1.00	0.67	0.80	0.83
Texture	0.88	0.84	0.86	0.86
Combined	0.88	0.87	0.88	0.88

Discussion

Overview

The aim of our study was to develop a computer vision–assisted AI method to extract, analyze, and recognize the facial features of patients with PD. Our results show that the accuracy of the computer vision–assisted AI method is more than 80%, and it has the advantages of timeliness, convenience, and low cost. Our results indicated that the computer vision–assisted AI method can play an important role in clinical practice for PD. First, it can analyze and summarize the facial features of patients with PD to form computer-recognizable digital information. And it can explore the representative and quantitative features of the facial features of patients with PD, which may help diagnose the disease and evaluate the effect of treatment. Second, through this method, large-scale, fast, convenient, and low-cost screening of PD in the population could be realized, helping patients who suspect PD go to the hospital as early as possible. Third, it is helpful for telemedicine and contactless medical treatment, which is of great significance during the COVID-19 pandemic. In the future, we will further improve the application of the computer vision–assisted AI method in the screening process for PD.

Principal Findings

In this study, we used facial features combined with machine learning algorithms to distinguish PD from healthy people. First, our study confirms that the facial expressions of PD patients are different from those of healthy people, which is consistent with other studies. Second, different from the use of contact sensors to analyze facial mimicry in PD patients to achieve the purpose of identification [25], our study is contact-free and has a more profound significance for human-computer interaction. At the same time, our method is more streamlined, and the processing time is short so that we can accelerate the process of diagnosis and make it more efficient. Third, the results indicated that the combination of texture features and geometric features of PD patients could be helpful for the diagnosis of PD. Texture features show better discrimination than geometric features. What is more, the abnormal facial expression around the eyes is more pronounced than around on mouth, while the degree of abnormal facial expression was not correlated with clinically evaluated variables of PD. Our results suggest that the facial features of PD patients may be a characteristic symptom that is solely related to the disease and is not affected by the motor symptoms, nonmotor symptoms, medications, and the course of the disease. Of course, the possibility for this may come from the statistical deviation caused by an insufficient sample size and sampling error, which cannot be ignored. Further studies are necessary to confirm these findings in other neurology centers.

Limitations

There were several limitations in our study. First, our study retrospectively enrolled treated PD patients. It is crucial to conduct prospective studies to enroll patients with de novo PD to verify the accuracy of the video of the facial feature recognition–based AI model to assist with PD diagnosis. Second, we collected videos using a single 2D video camera, and the amount of data was not sufficient. Although we were unable to capture the 3D facial expressions of the patients, 3D cameras can be used to capture more information in future research. Third, as all videos in the training data sets were classified as either PD or control, with no enrolled patients with parkinsonism, interpretation of these results cannot be extrapolated to other contexts. Facial tremor was not included as a parameter in our study for the following reasons. On the one hand, it requires expensive equipment to capture tremor information without physical contact. On the other hand, incorporating an indicator of tremor is time-consuming, because the correlation and frequency distribution characteristics between consecutive frames need to be calculated, which will greatly increase the amount of calculation required by the

system. What is more, tremor greatly interferes with facial expression recognition. Although we did not involve the analysis of facial tremor in this study, we can measure it using special sensors in the future.

Comparison With Prior Work

There was a similar study using facial features to diagnosis PD [27]. We verified the method they reported in our dataset. The results are compared with our work in Table 3. Our AI model showed better accuracy in the diagnosis of PD. Moreover, their study did not provide information on clinical characteristics of patients with PD.

Table 3. Comparison of results with those of prior work.

Work	Algorithm	Precision	Recall	F1 value
Jin et al [27]	SVM[a]	0.78	0.7	0.74
	RF[b]	0.6	0.9	0.72
Our method	SVM	0.8	0.87	0.84
	RF	0.88	0.87	0.88

[a]SVM: support vector machine.

[b]RF: random forest.

Conclusions

In summary, we have verified that facial feature information is effective for distinguishing PD from MC participants. All the geometric features, texture features, and combined features had good performance. Facial features play a role in the auxiliary diagnosis of PD. A markerless 2D video, facial feature recognition–based, AI model can provide a valuable tool to assist with PD diagnosis and the potential of realizing remote monitoring especially during the COVID-19 pandemic.

Acknowledgments

This study was supported by the Shanghai Health and Family Planning Commission Foundation (201940021) and Shanghai Pujiang Program (2020PJD032).

YZ is also a co-corresponding author for this work and can be contacted at zhangyu06@xinhuamed.com.cn

Authors' Contributions

XH analyzed and interpreted the data, drafted the manuscript, and critically revised the manuscript. YW, XW, JZ, and XZ enrolled the patients. YZ and JS critically revised the manuscript and supervised the study.

Conflicts of Interest

None declared.

References

1. Jankovic J. Parkinson disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008 Apr 01;79(4):368-376. [doi: 10.1136/jnnp.2007.131045] [Medline: 18344392]
2. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease. Neurology 2016 Jan 13;86(6):566-576. [doi: 10.1212/wnl.0000000000002350]
3. Cabestany C, López P, Sama A, Moreno JM, Bayes A, Rodriguez-Molinero A. REMPARK: When AI and technology meet Parkinson Disease assessment. 2013 Presented at: 20th International Conference Mixed Design of Integrated Circuits and Systems-MIXDES; June 20-22, 2013; Gdynia, Poland.
4. Agarwal A, Chandrayan S, Sahu SS. Prediction of Parkinson disease using speech signal with Extreme Learning Machine. 2016 Presented at: International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT); March 3-5, 2016; Chennai, India. [doi: 10.1109/iceeot.2016.7755419]
5. Pereira R, Weber SA, Hook C, Rosa GH, Papa JP. Deep learning-aided Parkinson disease diagnosis from handwritten dynamics. 2016 Presented at: 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); October 4-7, 2016; Sao Paulo, Brazil. [doi: 10.1109/sibgrapi.2016.054]
6. Hemmesch AR, Tickle-Degnen L, Zebrowitz LA. The influence of facial masking and sex on older adults' impressions of individuals with Parkinson disease. Psychol Aging 2009 Sep;24(3):542-549 [FREE Full text] [doi: 10.1037/a0016105] [Medline: 19739910]
7. Simons G, Pasqualini MCS, Reddy V, Wood J. Emotional and nonemotional facial expressions in people with Parkinson disease. J Int Neuropsychol Soc 2004 Jul 01;10(4):521-535. [doi: 10.1017/s135561770410413x]

8. Bowers D, Miller K, Bosch W, Gokcay D, Pedraza O, Springer U, et al. Faces of emotion in Parkinson disease: Micro-expressivity and bradykinesia during voluntary facial expressions. J. Inter. Neuropsych. Soc 2006 Oct 25;12(06):765. [doi: 10.1017/s135561770606111x]
9. Zhao Q, Okada K, Rosenbaum K, Kehoe L, Zand DJ, Sze R, et al. Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA. Med Image Anal 2014 Jul;18(5):699-710. [doi: 10.1016/j.media.2014.04.002] [Medline: 24835178]
10. Basel-Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz I, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. Clin Genet 2016 May 25;89(5):557-563. [doi: 10.1111/cge.12716] [Medline: 26663098]
11. Hamm J, Kohler CG, Gur RC, Verma R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. J Neurosci Methods 2011 Sep 15;200(2):237-256 [FREE Full text] [doi: 10.1016/j.jneumeth.2011.06.023] [Medline: 21741407]
12. Madeley P, Ellis AW, Mindham RHS. Facial expressions and Parkinson disease. Behavioural Neurology 1995;8(2):115-119. [doi: 10.1155/1995/750425]
13. Lee D, Park H, Yoo CD. Face alignment using cascade gaussian process regression trees. 2015 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 7-12, 2015; Boston, MA. [doi: 10.1109/cvpr.2015.7299048]
14. Dalal N, Triggs B. Histograms of oriented gradients for human detection. 2005 Presented at: IEEE computer society conference on computer vision and pattern recognition (CVPR'05); June 20-25, 2005; San Diego, CA. [doi: 10.1109/cvpr.2005.177]
15. Déniz O, Bueno G, Salido J, De la Torre F. Face recognition using Histograms of Oriented Gradients. Pattern Recognition Letters 2011 Sep;32(12):1598-1603. [doi: 10.1016/j.patrec.2011.01.004]
16. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Machine Intell 2002 Jul;24(7):971-987. [doi: 10.1109/tpami.2002.1017623]
17. Ahonen T, Hadid A, Pietikäinen M. Face Recognition with Local Binary Patterns. In: Pajdla T, Matas J, editors. Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science, vol 3021. Berlin, Heidelberg: Springer; 2004:469-481.
18. Wang T, Han X, Yan S. An HOG-LBP human detector with partial occlusion handling. 2009 Presented at: 12th international conference on computer vision; September 29-October 2, 2009; Kyoto, Japan p. 32-39. [doi: 10.1109/iccv.2009.5459207]
19. Noble WS. What is a support vector machine? Nat Biotechnol 2006 Dec;24(12):1565-1567. [doi: 10.1038/nbt1206-1565] [Medline: 17160063]
20. Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines. Microsoft. 1998 Apr. URL: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/ [accessed 2021-10-13]
21. Keller JM, Gray MR, Givens JA. A fuzzy K-nearest neighbor algorithm. IEEE Trans. Syst., Man, Cybern 1985 Jul;SMC-15(4):580-585. [doi: 10.1109/tsmc.1985.6313426]
22. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 1967 Jan;13(1):21-27. [doi: 10.1109/tit.1967.1053964]
23. Zhou ZH, Feng J. Deep forest. National Science Review 2019 Jan:74-86. [doi: 10.1093/nsr/nwy108]
24. Breiman L. Random forests. Machine learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
25. Hsieh W. Machine learning methods in the environmental sciences: Neural networks and kernels. Cambridge, MA: Cambridge University Press; Jul 30, 2009.
26. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
27. Jin B, Qu Y, Zhang L, Gao Z. Diagnosing Parkinson disease through facial expression recognition: video analysis. J Med Internet Res 2020 Jul 10;22(7):e18697 [FREE Full text] [doi: 10.2196/18697] [Medline: 32673247]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
AUC: area under the curve
HAM-A: Hamilton Anxiety Scale
HOG: histogram of oriented gradients
H-Y: Hoehn-Yahr
KNN: k-nearest neighbor
LBP: local binary pattern
LEDD: levodopa equivalent daily dose
MC: matched controls
MMSE: Mini-Mental State Examination
NMSS: Non-Motor Symptoms Scale

OR: odds ratio
PD: Parkinson disease
PDQ-39: Parkinson Disease Questionaire-39
QUIP: Questionnaire for Impulsive-Compulsive Disorders in Parkinson disease
RBDQ-HK: REM Sleep Behavior Disorder Questionnaire Hong Kong
RF: random forest
SVM: support vector machine
UPDRS: Unified Parkinson Disease Rating Scale

Edited by R Kukafka; submitted 12.04.21; peer-reviewed by Z Wang, S Jameela; comments to author 21.07.21; revised version received 23.08.21; accepted 22.09.21; published 19.11.21.

Please cite as:
Hou X, Zhang Y, Wang Y, Wang X, Zhao J, Zhu X, Su J
A Markerless 2D Video, Facial Feature Recognition–Based, Artificial Intelligence Model to Assist With Screening for Parkinson Disease: Development and Usability Study
J Med Internet Res 2021;23(11):e29554
URL: https://www.jmir.org/2021/11/e29554
doi:10.2196/29554
PMID:34806994

Original Paper

Artificial Intelligence for Skin Cancer Detection: Scoping Review

Abdulrahman Takiddin[1,2], BSc, MSc; Jens Schneider[2], BSc, MSc, PhD; Yin Yang[2], BSc, MSc, PhD; Alaa Abd-Alrazaq[2], BSc, MSc, PhD; Mowafa Househ[2], BSc, MSc, PhD

[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, United States
[2]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Abdulrahman Takiddin, BSc, MSc
Department of Electrical and Computer Engineering
Texas A&M University
188 Bizzell St
College Station, TX, 77843
United States
Phone: 974 44230425
Email: abdulrahman.takiddin@tamu.edu

Abstract

Background: Skin cancer is the most common cancer type affecting humans. Traditional skin cancer diagnosis methods are costly, require a professional physician, and take time. Hence, to aid in diagnosing skin cancer, artificial intelligence (AI) tools are being used, including shallow and deep machine learning–based methodologies that are trained to detect and classify skin cancer using computer algorithms and deep neural networks.

Objective: The aim of this study was to identify and group the different types of AI-based technologies used to detect and classify skin cancer. The study also examined the reliability of the selected papers by studying the correlation between the data set size and the number of diagnostic classes with the performance metrics used to evaluate the models.

Methods: We conducted a systematic search for papers using Institute of Electrical and Electronics Engineers (IEEE) Xplore, Association for Computing Machinery Digital Library (ACM DL), and Ovid MEDLINE databases following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidelines. The studies included in this scoping review had to fulfill several selection criteria: being specifically about skin cancer, detecting or classifying skin cancer, and using AI technologies. Study selection and data extraction were independently conducted by two reviewers. Extracted data were narratively synthesized, where studies were grouped based on the diagnostic AI techniques and their evaluation metrics.

Results: We retrieved 906 papers from the 3 databases, of which 53 were eligible for this review. Shallow AI-based techniques were used in 14 studies, and deep AI-based techniques were used in 39 studies. The studies used up to 11 evaluation metrics to assess the proposed models, where 39 studies used accuracy as the primary evaluation metric. Overall, studies that used smaller data sets reported higher accuracy.

Conclusions: This paper examined multiple AI-based skin cancer detection models. However, a direct comparison between methods was hindered by the varied use of different evaluation metrics and image types. Performance scores were affected by factors such as data set size, number of diagnostic classes, and techniques. Hence, the reliability of shallow and deep models with higher accuracy scores was questionable since they were trained and tested on relatively small data sets of a few diagnostic classes.

(J Med Internet Res 2021;23(11):e22934) doi:10.2196/22934

KEYWORDS

artificial intelligence; skin cancer; skin lesion; machine learning; deep neural networks

Introduction

Background

Skin cancer is the most common cancer type that affects humans [1]. Melanoma and nonmelanoma are the two main types of skin cancer [2]. Nonmelanoma is of lesser concern since it usually can be cured by surgery and is nonlethal. Melanoma, however, is the most dangerous skin cancer type, with a high mortality rate, although it represents less than 5% of all skin cancer cases [1]. The World Health Organization (WHO) estimated 132,000 yearly melanoma cases globally. In 2015, 60,000 cases caused death [2].

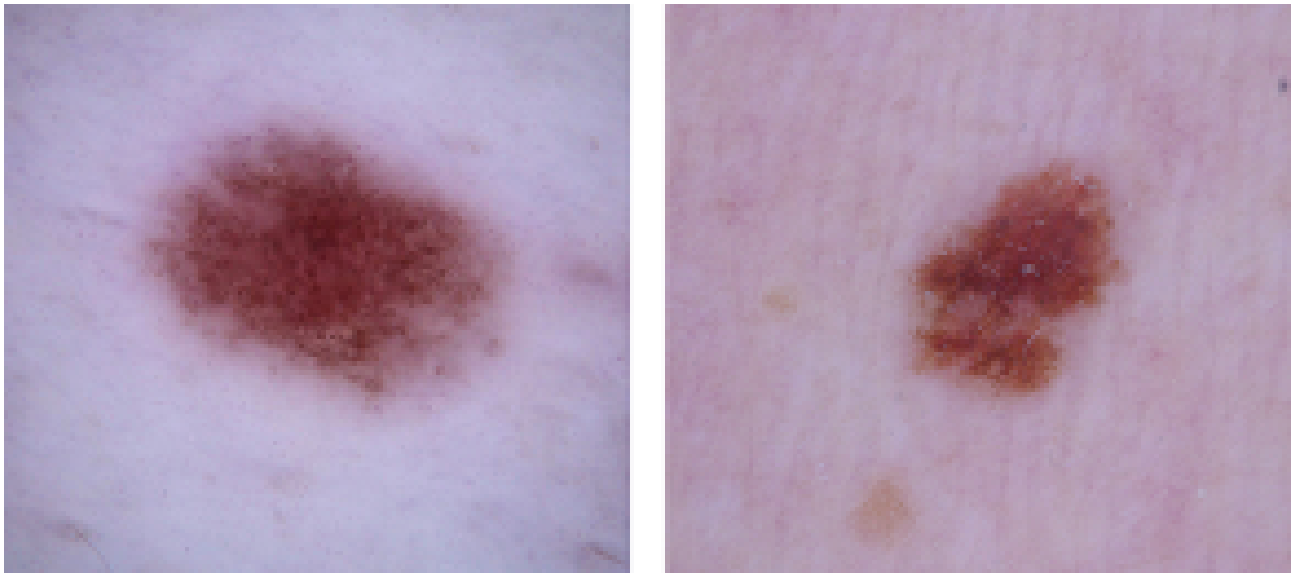
Traditional methods of early detection of skin cancer include skin self-examination and skin clinical examination (screening) [3]. However, skin self-examination, where the patient or a family member notices a lesion, is a random method as people might overreact or underact. In addition, clinical examination using expensive, specialized medical tools, such as a dermoscope, microspectroscopy, and laser-based tools, requires training, effort to operate, time, and regular follow-ups [4]. Thus, patients have started using mobile technologies, such as smartphones, to share images with their doctors to get faster diagnoses. However, sharing images over the internet may compromise privacy. Worse yet, the image quality may not be sufficient, which may lead to inaccurate diagnoses. With evolvement, artificial intelligence (AI), which is the human-like intelligence exhibited by trained machines [5], has become so pervasive that most humans interact with AI-based tools daily, which assists physicians in decision making and decreases the decision variations among physicians. It is worth mentioning that even with the presence of such AI technologies, the role of an expert dermatologist is vital for diagnosis and treatment.

The focus of this review is on the use of AI as a tool that helps in the process of skin cancer diagnostics. Herein, AI-based skin cancer diagnostic tools use either shallow or deep AI methodologies. Both involve customizing computer algorithms through a process called training to learn from data formed by predefined features. The difference is that shallow methods tend to not use multilayer neural networks at all or use such networks limited to a minimum of layers [6]. In contrast, deep methodologies involve training large, deep multilayer neural networks with many hidden layers, typically ranging from dozens to hundreds [7].

Research Problem

Detecting skin cancer can be challenging, time consuming, and relatively expensive [4]. For example, Figure 1 shows two lesions that superficially seem identical [8]. However, the left image is of a normal benign lesion, whereas the right image shows a melanoma lesion. As AI technologies are becoming smarter and faster [5], it is hardly surprising that they are being used to assist in diagnosing skin cancer and suggesting courses of action. This is due to the fact that AI-based methods are considered to be relatively cheap, easy to use, and accessible [5]. Thus, they offer the potential to overcome the issues inherent in the aforementioned existing skin cancer detection methods. However, as the literature on the medical use of AI quickly grows and continues to report findings using incompatible performance metrics, direct comparison between prior work becomes more challenging and threatens to hamper future research. This study seeks to address this issue by performing a rigorous and transparent review of the existing literature. We aim to answer the research question, *What are the existing AI-based tools that are used to detect and classify skin cancer?*

Figure 1. Similarity of normal lesion (left) and melanoma (right).



Methods

This scoping review analyzes papers from different online databases. We defined strict inclusion and exclusion criteria to decide which papers to include. We then grouped the papers by the methodology used and analyzed the ground covered in the papers. Finally, we identified gaps in the literature and discussed how these gaps can be filled by future work. We developed a protocol before commencing the review. To ensure that this scoping review is transparent and replicable, we followed the Preferred Reporting Items for Systematic Reviews and

Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) instructions and guidelines [9].

Search Strategy

We conducted a systematic search on July 15, 2020. We identified articles from Institute of Electrical and Electronics Engineers (IEEE) Xplore, Association for Computing Machinery Digital Library (ACM DL), and Ovid MEDLINE databases. The terms used for searching the bibliographic databases were identified based on the target population (eg, "skin neoplasms," "skin cancer," "skin lesion"), intervention (eg, "artificial intelligence," "machine learning," "deep learning"), and outcome ("diagnosis," "screening," "detection," "classification"). We derived the search terms from previous literature studies and reviews. For practical reasons, we did not conduct backward or forward reference list checking, and we also did not contact experts. Multimedia Appendix 1 shows the search strategy used for searching Ovid MEDLINE, where "skin neoplasms," "artificial intelligence," "machine learning," and "deep learning" were used as MESH terms. Multimedia Appendix 1 also shows the search query for IEEE Xplore and ACM DL.

Study Eligibility Criteria

We included studies fulfilling the following criteria:

- Studies published between January 1, 2009, and July 15, 2020.
- Studies written in English.
- Population: studies discussing only skin cancer. Studies discussing other diseases or forms of cancer were excluded.
- Intervention: studies discussing only AI-based applications. Studies that discussed skin cancer–related applications or systems, including theoretical, statistical, or mathematical approaches, were excluded.
- Studies discussing the specific use of AI for detecting, classifying, or diagnosing skin cancer. Studies discussing only the general use of AI in a clinical setting were excluded.
- Studies proposing a new AI-based method. Case studies, surveys, review or response papers, or papers that reviewed, assessed, analyzed, evaluated, or compared existing methods were excluded.

No restrictions on the country of publication, study design, comparator, or outcomes were enforced.

Study Selection

Authors Abdulrahman Takiddin (AT) and Alaa Abd-Alrazaq (AA) independently screened the titles and abstracts of all retrieved studies. Following the written protocol, they independently read the full texts of the papers included in this study after reading their titles and abstracts. Any disagreements between both reviewers were resolved by discussion. We assessed the intercoder agreement by calculating the Cohen kappa (κ), which was 0.86 and 0.93 for screening titles and abstracts and for reading full texts, respectively, indicating good agreement.

Data Extraction

For reliable and accurate data extraction from the included studies, a data extraction form was developed and piloted using eight included studies (Multimedia Appendix 2). The data extraction process was independently conducted by AT and AA. Any disagreements were resolved by discussion with good intercoder agreement (Cohen κ=0.88) between the reviewers.

Data Synthesis

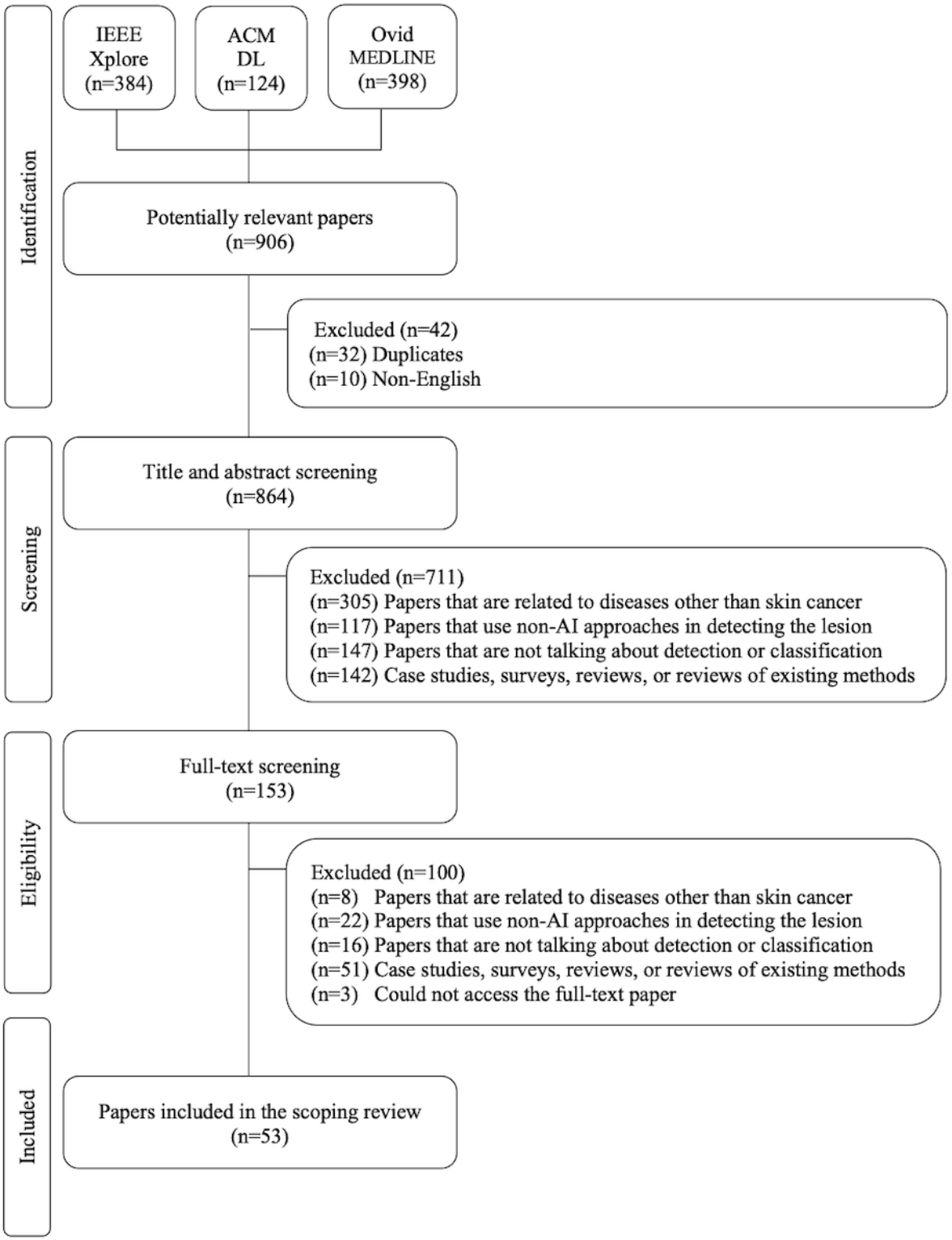
A narrative approach was used to synthesize the extracted data. Specifically, we first grouped the included studies by diagnostic techniques based on complexity. Then, we discussed the evaluation metrics used in each study. Next, we grouped the studies based on the used evaluation metrics. In addition, we took into consideration the used data set in terms of the number of images, types of images, and number of diseases (diagnostic classes) that the data set contained. We assessed the correlation between the accuracy score and the number of images and diagnostic classes of the data set.

Results

Search Results

After searching the 3 online databases, we retrieved a total of 906 studies. We then started excluding papers in three phases. As shown in Figure 2, in the first phase, "identification," we excluded 42 papers. In the second phase, "screening," we excluded 711 papers. In the last phase, "eligibility," we included 153 papers for a full-text review. After reviewing the full text of the papers, we excluded 100 papers. The specific reasons behind excluding the papers in each phase are mentioned in Figure 2. Hence, the total number of included papers in this scoping review was 53.

Figure 2. PRISMA approach. ACM DL: Association for Computing Machinery Digital Library; AI: artificial intelligence; IEEE: Institute of Electrical and Electronics Engineers; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Study Characteristics

Table 1 summarizes the characteristics of the selected studies. Figure 3 shows the number of papers published per year: 4 of 53 studies (7.6%) were published before 2016 [10-13], 26 studies (49.1%) were published in 2016, 2017, and 2018 [14-39], and 23 studies (43.4%) were published in 2019 and 2020 [40-62]. Although our selection criteria included papers published between 2009 and July 2020, the oldest published paper included after the full-text review was published in 2011. We observed that the number of papers sharply increased in 2018 and 2019.

Table 1. Study characteristics (N=53).

Characteristics	n (%)
Publication year	
Before 2016	4 (7.5)
2016-2018	26 (49.1)
2019-2020	23 (43.4)
Country of publication	
The United States	9 (16.9)
China	6 (11.3)
India	5 (9.4)
Poland	3 (5.7)
New Zealand	2 (3.8)
Austria	2 (3.8)
Germany	2 (3.8)
Bangladesh	2 (3.8)
Indonesia	2 (3.8)
Pakistan	2 (3.8)
Turkey	2 (3.8)
France	1 (1.9)
Russia	1 (1.9)
The United Kingdom	1 (1.9)
Hong Kong	1 (1.9)
Iran	1 (1.9)
Korea	1 (1.9)
Philippines	1 (1.9)
Lebanon	1 (1.9)
Saudi Arabia	1 (1.9)
Singapore	1 (1.9)
Thailand	1 (1.9)
Australia	1 (1.9)
Canada	1 (1.9)
Egypt	1 (1.9)
Nigeria	1 (1.9)
South Africa	1 (1.9)
Publication type	
Conference proceedings	31 (58.5)
Journals	22 (41.5)

Figure 3. Number of published papers by year.

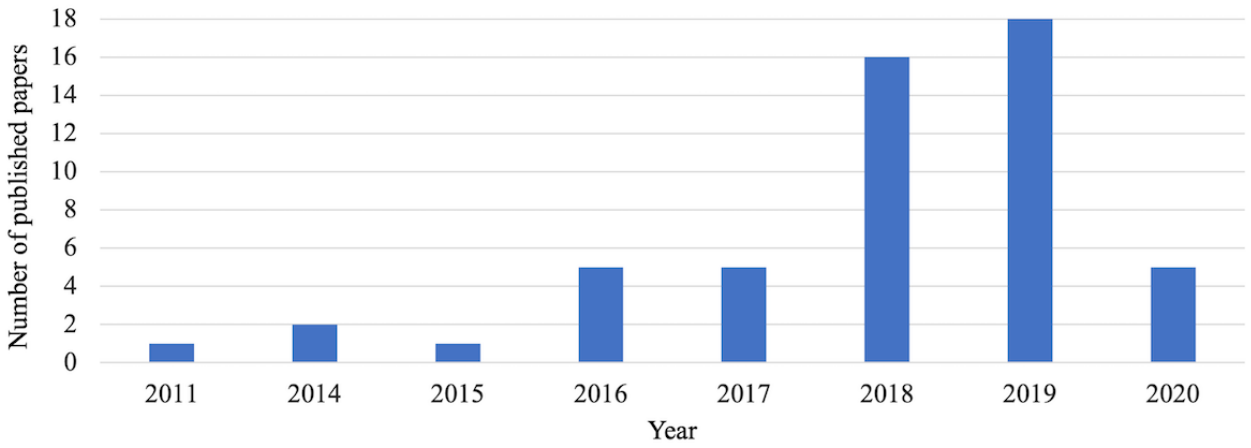
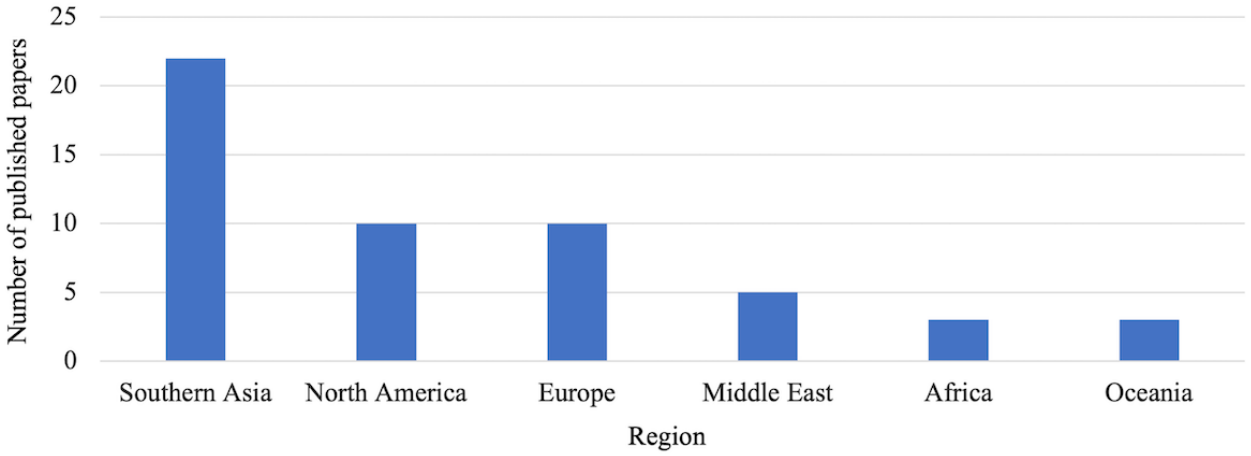


Figure 4 shows the region of publication of the included studies. The studies included were published in different parts of the world. In Southern Asia, 22 studies (41.5%) were conducted in China, India, Bangladesh, Indonesia, Pakistan, Singapore, South Korea, and Thailand; 10 studies (18.9%) were conducted in North America, specifically the United States and Canada; 10 studies were conducted in Europe, including Austria, Poland, Germany, France, the United Kingdom, and Russia; 5 studies (9.4%) were conducted in the Middle East, including Lebanon, Turkey, Iran, and Saudi Arabia; 3 studies (5.7%) were conducted in Africa, specifically Egypt, South Africa, and Nigeria; and in Oceania, 3 studies were concluded in New Zealand and Australia.

The selected studies were either published in conference proceedings or journals: 31 of 53 studies (58.5%) were published in conference proceedings, and the rest of the papers (22/53, 41.5%) were published in journals. Multimedia Appendix 3 displays the characteristics of each included study.

Figure 4. Number of published papers by region.



Data Characteristics

Table 2 summarizes the characteristics of the used data in the selected studies. The studies used different sizes of data sets to train their models. The average number of used images in the selected studies was around 7800. The lowest number of images used was 40 [24], whereas the highest number of images used was 129,450 [23]. We categorized these data set sizes into three groups, depending on the number of images used. The first category contained small data sets that had fewer than 1000 images (21/53, 39.6%). The second category used medium-size data sets consisting of 1000-10,000 images (25/53, 47.2%). The last category contained large data sets that included more than 10,000 images (7/53, 13.2%).

We divided the papers into two groups based on the classification type. We found that more than half of the papers (31/53, 58.5%) built models to classify whether the lesion was benign or malignant (two-class/binary classification). The rest of the papers (22/53, 41.5%) presented models in which skin lesions were classified using three or more diagnostic classes (multiclass classification). Figure 5 shows the number of papers using different diagnostic classes. In the multiclass classification, 8 studies used 3 diagnostic classes, 1 study used 4 classes, 2 studies used 5 classes, 10 studies used 7 classes, and 1 study used 9 classes. The benign classes included benign keratosis, melanocytic nevus, and dermatofibroma. The malignant classes included melanoma and basal cell carcinoma. Other lesions, such as vascular lesions, actinic keratosis, genodermatosis, and tumors, could be either benign or malignant.

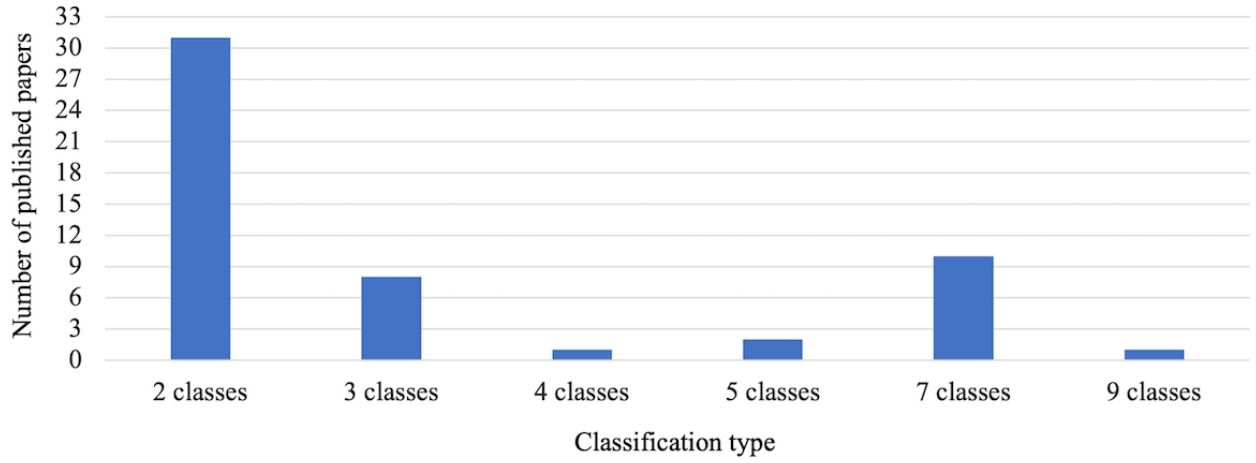
Table 2. Data and deployment characteristics (N=53).

Characteristics	n (%)
Data set size	
Small	21 (39.6)
Medium	25 (47.1)
Large	7 (13.2)
Classification type	
2 classes	31 (58.5)
3 classes	8 (15.1)
4 classes	1 (1.9)
5 classes	2 (3.8)
7 classes	10 (18.9)
9 classes	1 (1.9)
Image type	
Dermoscopic	43 (81.1)
Clinical	5 (9.4)
High quality	4 (7.5)
Spectroscopic	1 (1.9)
Deployment	
Development	45 (84.9)
System	3 (5.7)
Web application	3 (5.7)
Mobile application	2 (3.8)

With regard to the type of images used to train, test, and validate the models, 43 of 53 studies (81.1%) used dermoscopic images; 5 studies (9.4%) used clinical images that were taken using a normal camera; and 4 studies (7.5%) used high-quality images that were taken with a professional camera. The remaining study used spectroscopic images requiring a specialized system taking images of a lesion from three different spots using polarized and unpolarized light.

The majority of the studies (45/53, 84.9%) presented technologies that are still in the development phase. The rest of the studies (8/53, 15.1%) have been deployed into a usable form: 3 studies developed a health care system, 3 studies deployed the model into a mobile application, and 2 studies transferred the model into a web application. Multimedia Appendix 4 displays the data and deployment characteristics of each included study.

Figure 5. Number of published papers by number of diagnostic classes used.



Diagnostic Techniques

We categorized the papers into two groups based on the AI technique used in detecting and classifying skin cancer. The groups were *shallow* techniques and *deep* techniques. These two groups differed mainly in the complexity of the AI architecture underlying the model. *Shallow* techniques use either simple machine learning algorithms, such as a support vector machine (SVM), or only a couple of layers of neural networks [63]. If, in contrast, the AI architecture is a neural network that consists of at least three layers, it is categorized as a *deep* technique [19]. It turns out that around a quarter of the studies (14/53, 26.4%) used shallow techniques, while the rest (39/53, 73.6%) used deep techniques. Within each of the groups, studies may have used different models or algorithms, and some studies proposed multiple methods or provided testing data using multiple methods. In this study, we only considered the model that had the best-reported performance in each paper.

As shown in Table 3, most studies that used *shallow* techniques adopted an SVM (9/14, 64.3%), which is a common two-class classifier that uses a hyperplane as a decision boundary [6]. The rest of the studies (5/14, 35.7%) adopted the naive Bayes (NB) algorithm (1/14, 7.1%), which is a probabilistic classifier that assumes conditional independence among the features [6]; logistic regression (LR; 1/14), which uses probability for prediction; k-nearest neighbors (kNNs; 1/14), which classify a sample based on samples close to it; and random forests (RFs; 1/14), which classify using decision trees [6]. A hybrid model (1/14) classified images through multiple iteratives using Adaboost and an SVM.

Table 3. Techniques used in included studies using shallow techniques (N=14).

Model	n (%)	Reference
SVM[a]	9 (64.3)	[12,15,16,19,21,26,27,29,60]
NB[b]	1 (7.1)	[11]
LR[c]	1 (7.1)	[13]
kNN[d]	1 (7.1)	[25]
RF[e]	1 (7.1)	[28]
Hybrid	1 (7.1)	[18]

[a]SVM: support vector machine.

[b]NB: naive Bayes.

[c]LR: logistic regression.

[d]kNN: k-nearest neighbor.

[e]RF: random forest.

The majority of the studies that used *deep* techniques (Table 4) adopted different types of convolutional neural networks (CNNs; 36/39, 92.3%), which assign importance to parts of images using ImageNet-pretrained architectures (18/39, 46.2%), including the residual network (ResNet), Inception, AlexNet, MobileNet, Visual Geometry Group (VGG), Xception, DenseNet, and GoogleNet. In addition, some of the CNN-based studies (11/39, 28.2%) built customized CNNs or ResNets. Moreover, some studies adopted different combinations of CNNs along with other models (hybrid models; 5/39, 12.8%), as well as using ensemble models (4/39, 10.3%); the remaining study (1/39, 2.6%) used the OpenCV library. Multimedia Appendix 5 provides further details regarding each of the models in terms of the method used, the number of layers (ranging from 1 to 121 layers), the method used for selecting the hyperparameters, and the performance of the proposed model with respect to other reported models within the study.

Table 4. Techniques used in included studies using deep techniques (N=39).

Model	n (%)	Reference
Pretrained CNNs[a]		
ResNet[b]	5 (12.8)	[22,41,49,50,54]
Inception	3 (7.7)	[23,42,56]
AlexNet	3 (7.7)	[34,35,39]
MobileNet	3 (7.7)	[45,51,55]
VGG[c]	2 (5.1)	[30,52]
Xception	1 (2.6)	[43]
DenseNet	1 (2.6)	[58]
Custom		
CNN	9 (23.1)	[14,24,40,47,53,57,59,61,62]
ResNet	2 (5.1)	[31,33]
Hybrid	5 (12.8)	[17,32,38,44,46]
Ensemble	4 (10.3)	[20,36,37,48]
OpenCV	1 (2.6)	[10]

[a]CNN: convolutional neural network.

[b]ResNet: residual network.

[c]VGG: Visual Geometry Group.

Evaluation Metrics

The studies included in this scoping review used different evaluation metrics to assess their proposed models. In the studies, the following five primary evaluation metrics were used to assess the built models: accuracy, sensitivity and specificity, positive predictive value (PPV) or precision, area under the curve (AUC), and F1-score. All five metrics ranged from 0% to 100%; the higher the score, the better the model performance. To compute the different evaluation metrics, the following types of samples were identified: First, true positives (TPs), which are malignant samples that the AI tool also detected as malignant; second, false positives (FPs), which are benign samples that the AI tool detected as malignant; third, true negatives (TNs), which are benign samples that were also detected as benign by the AI tool; and fourth, false negatives (FNs), which are malignant samples that were detected as benign by the AI tool. It is worth mentioning that more than half of the studies (33/53, 62.3%) reported multiple evaluation metrics, in addition to the primary metric.

Accuracy = (TP + TN)/(TP + TN + FP + FN), which implies how well the model detects the diagnostic classes, was reported in the majority of the papers (44/53, 83%). Sensitivity or recall = TP/(TP + FN), which is the probability of the model, given only malignant samples, to correctly diagnose them as malignant, was reported in 30 (56.6%) papers. Specificity = TN/(TN + FP), which determines the proportion of negative samples that are correctly detected, was reported in 24 (45.3%) papers. The PPV or precision = TP/(TP + FP) was reported in 13 (24.5%) papers. The AUC, which is the area of the receiver operating characteristic (ROC) curve and plots the TP against the FP, was reported in 11 (20.8%) papers. The F1-score, which is the harmonic mean of recall and precision, was reported in 9 (16.9%) papers. In addition, the dice coefficient = 4TP/(FN + 2TP + FP) was reported in 4 (7.5%) papers. The negative predictive value (NPV) = TN/(TN + FN) was reported in 2 (3.8%) papers. The Jaccard index = 2TP/(TP + FN + FP) was reported in 2 papers. The Cohen κ was also reported in 2 papers. Finally, the Youden index = sensitivity + specificity – 1 was reported in 1 (1.9%) paper.

Herein, we conducted our analysis of each paper based on the best-performing experiment in case multiple experiments were conducted. In addition, if multiple evaluation metrics were used, we used the primary evaluation metric score that was reported by the authors in the abstract or conclusion as the main focus of the paper or the used average score of each of the diagnostic classes for multiclass classification papers. Of the aforementioned metrics, accuracy, AUC, sensitivity and specificity, and the F1-score were used as the primary evaluation metrics. Around 73% (39/53) of the papers used accuracy as their primary evaluation metric to assess the trained models. The average accuracy value was 86.8%, with a maximum of 98.8% [60] and a minimum of 67% [10]. The AUC was reported in 9 studies, with an average score of 87.2%; the highest AUC score was 91.7% [41], whereas the lowest AUC score was 82.0% [26]. Sensitivity and specificity were used in 4 studies, and the F1-score was reported in 1 study. Multimedia Appendix 6 shows the data characteristics, used model, and evaluation scores for each included study (Table 5).

Table 5. Primary evaluation metrics and scores reported by included studies (N=53).

Score	Reference
Accuracy	
99%	[60]
98%	[21,27]
96%	[24]
95%	[17,22,61]
94%	[20,40]
93%	[16]
92%	[18]
91%	[51,52,62]
90%	[36,42,57]
89%	[11,43]
88%	[13,48]
87%	[25,49,53]
86%	[35,44,58]
84%	[34]
83%	[54,55]
81%	[14]
80%	[19]
77%	[28]
75%	[39,47,59]
72%	[23,56]
67%	[10]
AUC[a]	
92%	[41]
91%	[33,38]
89%	[32]
87%	[46]
85%	[37,50]
84%	[30]
82%	[26]
Sensitivity	
96%	[31]
90%	[15]
83%	[12]
77%	[29]
Specificity	
96%	[15]
90%	[12]
89%	[31]
70%	[29]
F1-score	
83%	[45]

[a]AUC: area under the curve.

Discussion

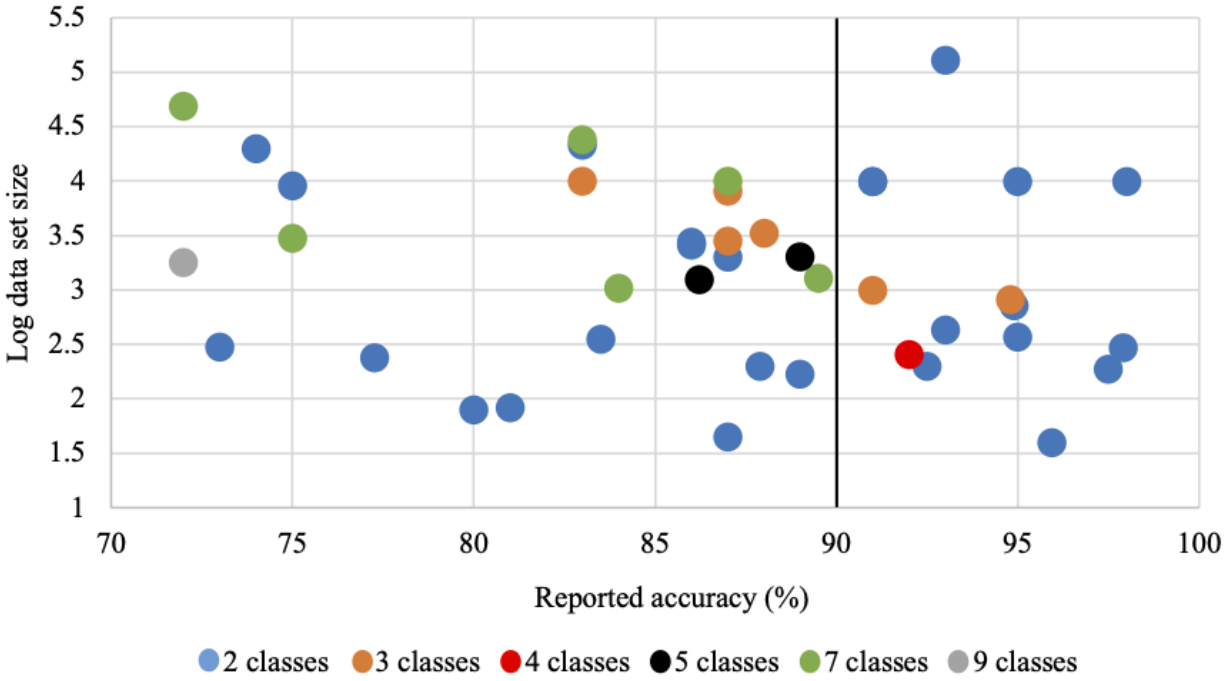
Main Findings

We studied multiple characteristic types for the 53 selected studies. First, we included the study characteristics. Most studies were published in 2019, the majority of the studies were published in Southern Asia, and most studies were published in journals. Second, we discussed the data characteristics. For training and testing, most of the studies used medium-size data sets, the majority of the studies built binary classifiers, and dermoscopic images were used the most. Third, we categorized the adopted AI models into shallow and deep. Most shallow models were SVM based, whereas most deep models were CNN-based neural networks. Generally, deep models were adopted more than shallow models. Fourth, we listed the evaluation metrics used along with the reported scores to assess the performance of the models. In total, 11 different evaluation metrics were used, where accuracy was the most commonly used metric, so we focused on accuracy.

Performance Factors

After analyzing the reported performance scores, we concluded that there is a correlation between the performance and the number of classes used. In addition, another factor that affects the performance is the data set size. Next, we study this hypothesis with respect to accuracy since most of the studies (39/53, 73.6%) used it as the primary evaluation metric, although it might not be the most fitted evaluation metric to assess such a task, especially in the case of imbalanced data. We believe that having a confusion matrix or the number of TPs, FPs, TNs, and FNs would avoid bias and give a clearer evaluation of how the model behaves with regard to each of the diagnostic classes. From the studies, the top accuracy scores were ~98% [21,27,60]. In studies leading to this accuracy, the authors built a two-class classification (benign vs malignant) model using data sets of 200, 356, and 200 images, respectively. The top 10 accuracy scores (99%-92%) also built two-class classifiers using an average of around 800 images. In addition, 26 studies built two-class classifiers with an average accuracy score of around 88% using an average data set size of around 1000 images, while 17 studies built multiclass classifiers with an average accuracy score of 85%; they used around 15,000 images on average. The second-lowest accuracy score was 72% [23], in which the authors developed a multiclass classifier using 9 different diagnostic classes and 129,450 images, which is the highest number of classes and the biggest data set size included in this study. Figure 6 plots the logarithmic data set size over accuracy, using colors to indicate the number of diagnostic classes. As can be seen, accuracy increases as the number of diagnostic classes and data set size decreases. Specifically, after the threshold of 90% in accuracy, we can see that the majority of the studies built two-class classifiers. The factors that might be behind such a pattern are further discussed next.

Figure 6. Effect of the number of diagnostic classes and data set size on accuracy.



Classification Type Factor

Binary classifiers tend to have better performance when compared to multiclass classifiers. This seems intuitively right since binary classifiers are less expressive. Instead of distinguishing between several classes, binary classifiers have "less to learn." To illustrate this point, let us compare limits on the probability of each class for a binary and a five-class classifier. For the five-class classifier, there must be at least one class with a probability of ≤20% (according to the *pigeonhole principle* [64]). Predicting this low probability class is, therefore, typically harder than in the case of a binary classifier, for which we know that there exists exactly (and, thus, at most) one class with a probability of ≤50%. Another way of looking at it is to

consider an algorithm that performs a random choice assuming perfectly balanced data. In the binary case, the error rate of this algorithm would be 50%, whereas for the five-class classifier, it increases to 80%, a 1.6-fold increase. The problem may be further exacerbated by imbalanced data, which often arises naturally due to differences in the prevalence rates of medical conditions. Therefore, it is also not surprising that binary classifiers work well, given less data for training, since the model may still be fed sufficient numbers of examples for each class.

Data Set Size Factor

However, what is surprising is that Figure 6 suggests that the performance increased with decreasing training data. To this end, we would like to note that the two methods with the best performance used shallow techniques that tend to be far less hungry for data than deep methods, since manual feature engineering is often part of the pipeline. Furthermore, Afifi et al [21] used clinical image data, which may be of superior quality. In addition, depending on the testing setup, it cannot be ruled out that methods relying on less data lack the generality of models that have been trained using large volumes of data. In such scenarios, the models would be closer to data retrieval machines due to overfitting than general detectors and classifiers. To fully assess apparent issues such as this, it is important not to rely on a single performance metric when reporting results. Especially, sensitivity and specificity can be as important as accuracy in this context since they model FN and FP rates. All considered, we would, therefore, like to reiterate our earlier statement that we believe it is important for any AI to undergo rigorous clinical studies and testing before being deployed in a clinical environment.

Technique Type Factor

With regard to the techniques described in the studies included in this review, deep and shallow models (regardless of the number of layers) have similar performances. For example, within the shallow models, the top five skin cancer detectors were built using an SVM with accuracy scores of 93%-99% using relatively small data sets. The SVM was the most commonly used method among the shallow models. Similarly, within the deep models, the top five CNN-based skin cancer detectors had 94%-96% accuracy using medium-size data sets. CNNs were also the most commonly used method among the deep models. Theoretically, deep neural networks tend to have better performance with regard to image classifications [65]. One reason is that shallow models are often limited to less expressive functional spaces when compared to deep networks. From a technical perspective, this may well explain their lower performance due to a lack of the ability to fully capture the complex nature of images during training. In contrast, deep networks and CNNs can learn features at multiple scales and complexity to provide fast diagnoses [66]. Therefore, they not only detect, select, and extract features from medical images but also contribute by enhancing and constructing new features from the medical images [67]. Such similarities and inconsistencies in the performances of the included studies are due to the diverse evaluation metrics used, the data set size, image types, and the number of diagnostic classes among the studies.

Publication Year

Based on the study characteristics, we noticed that the number of published papers has increased since 2016 and that most papers discuss the use of dermoscopic images, making it the most used image modality for the detection and classification of skin cancer. We believe that this is because the International Skin Imaging Collaboration (ISIC) competition started in 2016 [8], which offered several medical data sets of dermoscopic images that have ever since been used to build AI-based models. Most of these studies are still in the development stage, and we firmly believe that these models still need to be further validated and tested in hospitals. However, dermatologists and patients are beginning to adapt to the notion of relying on AI to diagnose skin cancer.

Practical and Research Implications

In this scoping review, we summarized the findings in the literature related to diagnosing skin cancer by using AI-based technology. We also categorized the papers included in this review based on the methodology used, the type of AI techniques, and their performance, and found the link between these aspects.

We noted that although all the papers included in this scoping review discuss the application and performance of a specific AI technology, the reporting is performed heterogeneously. A discussion of the relationship between using one specific AI technique and other aspects, such as data set size, or even a discussion of why the evaluation metric used is reasonable is normally not attempted. This, of course, potentially hampers research in this direction, as it becomes harder for future studies to provide a comprehensive comparison with the existing work that follows scientific rigor. This scoping review filled this gap by performing the necessary characterizations and analyses. This was achieved by grouping each of the used AI technologies into shallow and deep approaches, linking each type to the evaluation metrics used, listing and interpreting the number of diagnostic classes used in each study, and highlighting the dependency of performance on data set size and other factors. To the best of our knowledge, no similar work has been performed to fill this gap. In the Conclusion section, we will highlight our main findings.

Limitations

This scoping review examined papers that were published between January 2009 and July 2020, and any published study outside this time line was excluded, which may have excluded older AI-based methods. In addition, we examined papers written in English; other languages were not included, which may have led to the exclusion of some studies conducted in other parts of the world. Another limitation might be the gap between the time the research was performed and the time the work was submitted, which excluded published papers during that period. Although we applied all due diligence, a small residual chance of accidentally having overlooked papers in an academic database cannot be fully ruled out. In addition, although we tried to discuss all findings in the literature, it is

beyond the scope of this review to detail every single finding of the papers. Similarly, an investigation into data biases in the literature (imbalanced data with respect to diagnostic classes, patient ethnicity and skin color, gender, etc) is left as a direction for future studies.

Conclusions

The use of AI has high potential to facilitate the way skin cancer is diagnosed. Two main branches of AI are used to detect and classify skin cancer, namely shallow and deep techniques. However, the reliability of such AI tools is questionable since different data set sizes, image types, and number of diagnostic classes are being used and evaluated with different evaluation metrics. Accuracy is the metric used most as a primary evaluation metric but does not allow for independently assessing FN and FP rates. This study found that higher accuracy scores are reported when fewer diagnostic classes are included. Interestingly and counterintuitively, our analysis also suggests that higher accuracy scores are reported when smaller sample sizes are included, which may be due to factors such as the type of images and the techniques used. Furthermore, only independent, external validation using a large, diverse, and unbiased database is fit to demonstrate the generality and reliability of any AI technology prior to clinical deployment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search query.
[DOCX File , 16 KB - jmir_v23i11e22934_app1.docx]

Multimedia Appendix 2
Data extraction form.
[DOCX File , 14 KB - jmir_v23i11e22934_app2.docx]

Multimedia Appendix 3
Study characteristics.
[DOCX File , 20 KB - jmir_v23i11e22934_app3.docx]

Multimedia Appendix 4
Data and deployment characteristics.
[DOCX File , 21 KB - jmir_v23i11e22934_app4.docx]

Multimedia Appendix 5
Technical details.
[DOCX File , 32 KB - jmir_v23i11e22934_app5.docx]

Multimedia Appendix 6
Data, model, and evaluation.
[DOCX File , 32 KB - jmir_v23i11e22934_app6.docx]

References

1. Ray A, Gupta A, Al A. Skin lesion classification with deep convolutional neural network: process development and validation. JMIR Dermatol 2020 May 7;3(1):e18438. [doi: 10.2196/18438]
2. de Carvalho TM, Noels E, Wakkee M, Udrea A, Nijsten T. Development of smartphone apps for skin cancer risk assessment: progress and promise. JMIR Dermatol 2019 Jul 11;2(1):e13376. [doi: 10.2196/13376]
3. Loescher LJ, Janda M, Soyer HP, Shea K, Curiel-Lewandrowski C. Advances in skin cancer early detection and diagnosis. Semin Oncol Nurs 2013 Aug;29(3):170-181. [doi: 10.1016/j.soncn.2013.06.003] [Medline: 23958215]
4. Lieber CA, Majumder SK, Ellis DL, Billheimer DD, Mahadevan-Jansen A. In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy. Lasers Surg Med 2008 Sep;40(7):461-467 [FREE Full text] [doi: 10.1002/lsm.20653] [Medline: 18727020]
5. Murphy R. Introduction to AI Robotics. Cambridge, MA: MIT Press; 2019.
6. Marsland S. Machine Learning: An Algorithmic Perspective. Boca Raton, FL: CRC Press; 2011.
7. Mitra B, Craswell N. An Introduction to Neural Information Retrieval. Boston, MA: Now Foundations and Trends; 2018.
8. ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection. URL: https://challenge2018.isic-archive.com/ [accessed 2020-06-11]

9. 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]
10. Ramlakhan K, Shang Y. A mobile automated skin lesion classification system. 2011 Presented at: IEEE 23rd International Conference on Tools with Artificial Intelligence; 2011; Boca Raton, FL. [doi: 10.1109/ictai.2011.29]
11. Li L, Zhang Q, Ding Y, Jiang H, Thiers BH, Wang JZ. Automatic diagnosis of melanoma using machine learning methods on a spectroscopic system. BMC Med Imaging 2014 Oct 13;14(1):36 [FREE Full text] [doi: 10.1186/1471-2342-14-36] [Medline: 25311811]
12. Sabouri P, GholamHosseini H, Larsson T, Collins J. A cascade classifier for diagnosis of melanoma in clinical images. 2014 Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 26-30 Aug. 2014; Chicago, IL p. 6751. [doi: 10.1109/embc.2014.6945177]
13. Kaur R, Albano PP, Cole JG, Hagerty J, LeAnder RW, Moss RH, et al. Real-time supervised detection of pink areas in dermoscopic images of melanoma: importance of color shades, texture and location. Skin Res Technol 2015 Nov 22;21(4):466-473 [FREE Full text] [doi: 10.1111/srt.12216] [Medline: 25809473]
14. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using convolutional neural network. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016; Orlando, FL. [doi: 10.1109/embc.2016.7590963]
15. Jaworek-Korjakowska J. Computer-aided diagnosis of micro-malignant melanoma lesions applying support vector machines. Biomed Res Int 2016;2016:4381972 [FREE Full text] [doi: 10.1155/2016/4381972] [Medline: 27382567]
16. Jaworek-Korjakowska J, Kłeczek P. Automatic classification of specific melanocytic lesions using artificial intelligence. Biomed Res Int 2016;2016:8934242 [FREE Full text] [doi: 10.1155/2016/8934242] [Medline: 26885520]
17. Sabbaghi S, Aldeen M, Garnavi R. A deep bag-of-features model for the classification of melanomas in dermoscopy images. 2016 Presented at: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016; Orlando, FL p. 2016. [doi: 10.1109/embc.2016.7590962]
18. Premaladha J, Ravichandran KS. Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. J Med Syst 2016 Apr 12;40(4):96. [doi: 10.1007/s10916-016-0460-2] [Medline: 26872778]
19. Mustafa S, Dauda AB, Dauda M. Image processing and SVM classification for melanoma detection. 2017 Presented at: 2017 International Conference on Computing Networking and Informatics (ICCNI); 2017; Lagos, Nigeria p. 1-5. [doi: 10.1109/iccni.2017.8123777]
20. Xie F, Fan H, Li Y, Jiang Z, Meng R, Bovik A. Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE Trans Med Imaging 2017 Mar;36(3):849-858. [doi: 10.1109/tmi.2016.2633551]
21. Afifi S, GholamHosseini H, Sinha R. SVM classifier on chip for melanoma detection. 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017; Jeju, Korea (South). [doi: 10.1109/embc.2017.8036814]
22. Yu L, Chen H, Dou Q, Qin J, Heng P. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 2017 Apr;36(4):994-1004. [doi: 10.1109/tmi.2016.2642839]
23. 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 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
24. Mandache D, Dalimier E, Durkin J, Boceara C, Olivo-Marin J, Meas-Yedid V. Basal cell carcinoma detection in full field OCT images using convolutional neural networks. 2018 Presented at: IEEE 15th International Symposium on Biomedical Imaging (ISBI); 2018; Washington, DC. [doi: 10.1109/isbi.2018.8363689]
25. Linsangan N, Adtoon J. Skin cancer detection classification for moles using k-nearest neighbor algorithm. 2018 Presented at: 5th International Conference on Bioinformatics Research Applications; 2018; New York, NY. [doi: 10.1145/3309129.3309141]
26. Marchetti MA, Codella NC, Dusza SW, Gutman DA, Helba B, Kalloo A, International Skin Imaging Collaboration. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 2018 Feb;78(2):270-277.e1 [FREE Full text] [doi: 10.1016/j.jaad.2017.08.016] [Medline: 28969863]
27. Nasir M, Attique Khan M, Sharif M, Lali IU, Saba T, Iqbal T. An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. Microsc Res Tech 2018 Jun 21;81(6):528-543. [doi: 10.1002/jemt.23009] [Medline: 29464868]
28. Gautam D, Ahmed M, Meena YK, Ul Haq A. Machine learning-based diagnosis of melanoma using macro images. Int J Numer Method Biomed Eng 2018 May 20;34(5):e2953. [doi: 10.1002/cnm.2953] [Medline: 29266819]
29. Salem C, Azar D, Tokajian S. An image processing and genetic algorithm-based approach for the detection of melanoma in patients. Methods Inf Med 2018:231-286.
30. Yu C, Yang S, Kim W, Jung J, Chung K, Lee SW, et al. Acral melanoma detection using a convolutional neural network for dermoscopy images. PLoS ONE 2018 Mar 7;13(3):e0193321. [doi: 10.1371/journal.pone.0193321]
31. Putten EV, Kambod A, Kambod M. Deep residual neural networks for automated basal cell carcinoma detection. 2018 Presented at: IEEE EMBS International Conference on Biomedical Health Informatics (BHI); 2018; Las Vegas, NV. [doi: 10.1109/bhi.2018.8333437]

32. Yang X, Hangxing H, Wang L, Yeo S, Su Y, Zeng Z. Skin lesion analysis by multi-target deep neural networks. 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018; Honolulu, HI p. 2018. [doi: 10.1109/embc.2018.8512488]
33. Li Y, Shen L. Skin lesion analysis towards melanoma detection using deep learning network. Sensors (Basel) 2018 Feb 11;18(2):556 [FREE Full text] [doi: 10.3390/s18020556] [Medline: 29439500]
34. Kaymak S, Esmaili P, Serener A. Deep learning for two-step classification of malignant pigmented skin lesions. 2018 Presented at: 14th Symposium on Neural Networks and Applications (NEUREL); 2018; Belgrade, Serbia p. 1. [doi: 10.1109/neurel.2018.8587019]
35. Hameed N, Shabut A, Hossain M. Multi-class skin diseases classification using deep convolutional neural network and support vector machine. 2018 Presented at: 12th International Conference on Software, Knowledge, Information Management Applications (SKIMA); 2018; Phnom Penh, Cambodia. [doi: 10.1109/skima.2018.8631525]
36. Shahin A, Kamal A, Elattar M. Deep ensemble learning for skin lesion classification from dermoscopic images. 2018 Presented at: 9th Cairo International Biomedical Engineering Conference (CIBEC); 2018; Cairo, Egypt. [doi: 10.1109/cibec.2018.8641815]
37. Harangi B, Baran A, Hajdu A. Classification of skin lesions using an ensemble of deep neural networks. 2018 Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018; Honolulu, HI p. 2018. [doi: 10.1109/embc.2018.8512800]
38. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinge I. Skin lesion classification using hybrid deep neural networks. 2019 Presented at: ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019; Brighton, UK. [doi: 10.1109/icassp.2019.8683352]
39. Shihadeh J, Ansari A, Ozunfunmi T. Deep learning based image classification for remote medical diagnosis. 2018 Presented at: IEEE Global Humanitarian Technology Conference (GHTC); 2018; San Jose, CA. [doi: 10.1109/ghtc.2018.8601558]
40. Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. Int J Med Inform 2019 Apr;124:37-48. [doi: 10.1016/j.ijmedinf.2019.01.005] [Medline: 30784425]
41. Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification. IEEE Trans Med Imaging 2019 Sep;38(9):2092-2103. [doi: 10.1109/tmi.2019.2893944]
42. Demir A, Yilmaz F, Kose O. Early detection of skin cancer using deep learning architectures: ResNet-101 and Inception-v3. 2019 Presented at: Medical Technologies Congress (TIPTEKNO); 2019; Izmir, Turkey. [doi: 10.1109/tiptekno47231.2019.8972045]
43. Gavrilov D, Lazarenko L, Zakirov E. AI recognition in skin pathologies detection. 2019 Presented at: 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI); 2019; Belgrade, Serbia. [doi: 10.1109/ic-aiai48757.2019.00017]
44. Aggarwal A, Das N, Sreedevi I. Attention-guided deep convolutional neural networks for skin cancer classification. 2019 Presented at: Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA); 2019; Istanbul, Turkey. [doi: 10.1109/ipta.2019.8936100]
45. Liu Y. The application of deep learning on fast skin cancer diagnosis. 2019 Presented at: 2019 International Conference on Information Technology and Computer Application (ITCA); 2019; Guangzhou, China. [doi: 10.1109/itca49981.2019.00034]
46. Liang R, Wu Q, Yang X. Multi-pooling attention learning for melanoma recognition. 2019 Presented at: 2019 Digital Image Computing: Techniques and Applications (DICTA); 2019; Perth, WA, Australia. [doi: 10.1109/dicta47822.2019.8945868]
47. Dai X, Spasi I, Meyer B, Chapman S, Andres F. Machine learning on mobile: an on-device inference app for skin cancer detection. 2019 Presented at: Fourth International Conference on Fog and Mobile Edge Computing (FMEC); 2019; Rome, Italy p. A. [doi: 10.1109/fmec.2019.8795362]
48. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C. Fusing fine-tuned deep features for skin lesion classification. Comput Med Imaging Graph 2019 Jan;71:19-29. [doi: 10.1016/j.compmedimag.2018.10.007] [Medline: 30458354]
49. Wodzinski M, Skalski A, Witkowski A, Pellacani G, Ludzik J. Convolutional neural network approach to classify skin lesions using reflectance confocal microscopy. 2019 Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8856731]
50. Brinker TJ, Hekler A, Enk AH, von Kalle C. Enhanced classifier training to improve precision of a convolutional neural network to identify images of skin lesions. PLoS One 2019 Jun 24;14(6):e0218713 [FREE Full text] [doi: 10.1371/journal.pone.0218713] [Medline: 31233565]
51. Ech-Cherif A, Misbhauddin M, Ech-Cherif M. Deep neural network based mobile dermoscopy application for triaging skin cancer detection. 2019 Presented at: 2nd International Conference on Computer Applications Information Security (ICCAIS); 2019; Riyadh, Saudi Arabia. [doi: 10.1109/cais.2019.8769517]
52. Guha SR, Rafizul Haque SM. Convolutional neural network based skin lesion analysis for classifying melanoma. 2019 Presented at: International Conference on Sustainable Technologies for Industry 4.0 (STI); 2019; Dhaka, Bangladesh. [doi: 10.1109/sti47673.2019.9067979]

53. Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R. Depthwise separable convolutional neural network for skin lesion classification. 2019 Presented at: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT); 2019; Ajman, United Arab Emirates. [doi: 10.1109/isspit47144.2019.9001790]
54. Budhiman A, Suyanto S, Arifianto A. Melanoma cancer classification using ResNet with data augmentation. 2019 Presented at: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI); 2019; Yogyakarta, Indonesia. [doi: 10.1109/isriti48646.2019.9034624]
55. Sae-Lim W, Wettayaprasit W, Aiyarak P. Convolutional neural networks using MobileNet for skin lesion classification. 2019 Presented at: 16th International Joint Conference on Computer Science and Software Engineering (JCSSE); 2019; Chonburi, Thailand. [doi: 10.1109/jcsse.2019.8864155]
56. Purnama IKE, Hernanda AK, Ratna AAP, Nurtanio I, Hidayati AN, Purnomo MH, et al. Disease classification based on dermoscopic skin images using convolutional neural network in teledermatology system. 2019 Presented at: International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM); 2019; Surabaya, Indonesia. [doi: 10.1109/cenim48368.2019.8973303]
57. Hasan M, Barman SD, Islam S, Reza AW. Skin cancer detection using convolutional neural network. 2019 Presented at: 5th International Conference on ComputingArtificial Intelligence, New York, NY, USA; 2019; Bali, Indonesia. [doi: 10.1145/3330482.3330525]
58. Wei L, Ding K, Hu H. Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE Access 2020;8:99633-99647. [doi: 10.1109/access.2020.2997710]
59. Nasiri S, Helsper J, Jung M, Fathi M. DePicT Melanoma Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images. BMC Bioinformatics 2020 Mar 11;21(Suppl 2):84-13 [FREE Full text] [doi: 10.1186/s12859-020-3351-y] [Medline: 32164530]
60. Poovizhi S, Ganesh Babu TR. An efficient skin cancer diagnostic system using Bendlet Transform and support vector machine. An Acad Bras Cienc 2020;92(1):e20190554 [FREE Full text] [doi: 10.1590/0001-3765202020190554] [Medline: 32491128]
61. Adegun AA, Viriri S. Deep learning-based system for automatic melanoma detection. IEEE Access 2020;8:7160-7172. [doi: 10.1109/access.2019.2962812]
62. Sanketh RS, Madhu Bala M, Narendra Reddy PV, Phani Kumar GVS. Melanoma disease detection using convolutional neural networks. 2020 Presented at: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020; 2020; Madurai, India. [doi: 10.1109/iciccs48265.2020.9121075]
63. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.
64. Herstein I. Topics in Algebra. Waltham, MA: Blaisdell; 1964:90.
65. Yin X, Yang C, Pei W, Hao H. Shallow classification or deep learning: an experimental study. 2014 Presented at: 2014 22nd International Conference on Pattern Recognition; 2014; Stockholm, Sweden. [doi: 10.1109/icpr.2014.333]
66. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin cancer classification using convolutional neural networks: systematic review. J Med Internet Res 2018 Oct 17;20(10):e11936 [FREE Full text] [doi: 10.2196/11936] [Medline: 30333097]
67. Razzak MI, Zaib A. Deep learning for medical image processing: overview, challenges and the future. In: Classification in BioApps: Automation of Decision Making. Cham, Switzerland: Springer International; 2018:323.

Abbreviations

ACM DL: Association for Computing Machinery Digital Library
AI: artificial intelligence
AUC: area under the curve
CNN: convolutional neural network
FN: false negative
FP: false positive
IEEE: Institute of Electrical and Electronics Engineers
ISIC: International Skin Imaging Collaboration
kNN: k-nearest neighbor
LR: logistic regression
NB: naive Bayes
NPV: negative predictive value
PPV: positive predictive value
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
ResNet: residual network
RF: random forest
ROC: receiver operating characteristic

SVM: support vector machine
TN: true negative
TP: true positive
VGG: Visual Geometry Group

Edited by R Kukafka, G Eysenbach; submitted 27.07.20; peer-reviewed by J Makin, E Frontoni, JA Benítez-Andrades, S Shams, R Sutton; comments to author 17.11.20; revised version received 05.01.21; accepted 03.08.21; published 24.11.21.

Please cite as:
Takiddin A, Schneider J, Yang Y, Abd-Alrazaq A, Household M
Artificial Intelligence for Skin Cancer Detection: Scoping Review
J Med Internet Res 2021;23(11):e22934
URL: https://www.jmir.org/2021/11/e22934
doi:10.2196/22934
PMID:34821566

Original Paper

Patients' Perceptions Toward Human–Artificial Intelligence Interaction in Health Care: Experimental Study

Pouyan Esmaeilzadeh[1*], PhD; Tala Mirzaei[1*], PhD; Spurthy Dharanikota[1*], MBA

Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States
[*]all authors contributed equally

Corresponding Author:
Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics
College of Business
Florida International University
Modesto A Maidique Campus
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 305 348 330
Email: pesmaeil@fiu.edu

Abstract

Background: It is believed that artificial intelligence (AI) will be an integral part of health care services in the near future and will be incorporated into several aspects of clinical care such as prognosis, diagnostics, and care planning. Thus, many technology companies have invested in producing AI clinical applications. Patients are one of the most important beneficiaries who potentially interact with these technologies and applications; thus, patients' perceptions may affect the widespread use of clinical AI. Patients should be ensured that AI clinical applications will not harm them, and that they will instead benefit from using AI technology for health care purposes. Although human-AI interaction can enhance health care outcomes, possible dimensions of concerns and risks should be addressed before its integration with routine clinical care.

Objective: The main objective of this study was to examine how potential users (patients) perceive the benefits, risks, and use of AI clinical applications for their health care purposes and how their perceptions may be different if faced with three health care service encounter scenarios.

Methods: We designed a 2×3 experiment that crossed a type of health condition (ie, acute or chronic) with three different types of clinical encounters between patients and physicians (ie, AI clinical applications as substituting technology, AI clinical applications as augmenting technology, and no AI as a traditional in-person visit). We used an online survey to collect data from 634 individuals in the United States.

Results: The interactions between the types of health care service encounters and health conditions significantly influenced individuals' perceptions of privacy concerns, trust issues, communication barriers, concerns about transparency in regulatory standards, liability risks, benefits, and intention to use across the six scenarios. We found no significant differences among scenarios regarding perceptions of performance risk and social biases.

Conclusions: The results imply that incompatibility with instrumental, technical, ethical, or regulatory values can be a reason for rejecting AI applications in health care. Thus, there are still various risks associated with implementing AI applications in diagnostics and treatment recommendations for patients with both acute and chronic illnesses. The concerns are also evident if the AI applications are used as a recommendation system under physician experience, wisdom, and control. Prior to the widespread rollout of AI, more studies are needed to identify the challenges that may raise concerns for implementing and using AI applications. This study could provide researchers and managers with critical insights into the determinants of individuals' intention to use AI clinical applications. Regulatory agencies should establish normative standards and evaluation guidelines for implementing AI in health care in cooperation with health care institutions. Regular audits and ongoing monitoring and reporting systems can be used to continuously evaluate the safety, quality, transparency, and ethical factors of AI clinical applications.

(J Med Internet Res 2021;23(11):e25856) doi:10.2196/25856

KEYWORDS

AI clinical applications; collective intelligence; in-person examinations; perceived benefits; perceived risks

Introduction

Artificial Intelligence

Artificial intelligence (AI) generally refers to a computerized system (hardware or software) that can perform physical tasks and cognitive functions, solve various problems, or make decisions without explicit human instructions [1]. A range of techniques and applications are under the broad umbrella of AI, such as genetic algorithms, neural networks, machine learning, and pattern recognition [2]. AI is considered a frontline service technology (FST) in the literature [3]. FST infusion in various industries has emerged as a topic of interest in the past decade (eg, [4-8]). Gursoy et al [9] classified FST infusion under three main categories: (1) no FST, which refers to a technology-free encounter between consumers and frontline service providers; (2) augmenting FST, which refers to the technology as a human augmentation tool; and (3) substituting FST, which refers to the technology as a human substitution force. In augmenting FST, the technology can help to enhance human thinking, analysis, and behavior, and boost the ability to interact with other human actors, whereas in substituting FST, the technology substitutes a human actor and takes away the active role of humans in the service encounter. AI as an FST can augment or replace human tasks and activities within a wide range of industrial, intellectual, and social applications with potential impacts on productivity and performance. As nonhuman intelligence is programmed to complete specific tasks, AI can overcome some of humans' computationally intensive and intellectual limitations [10]. For example, AI could be a computer application that uses sophisticated algorithms to solve a business problem for managers. AI applications generate personalized recommendations to customers based on analysis of a huge data set. Thus, it is believed that AI could perform tasks better than the best humans and experts in any field [2].

AI technology, including algorithmic machine learning and autonomous decision-making, creates new opportunities for continued innovation in different industries, including finance, health care, manufacturing, retail, supply chain, logistics, and utilities [11]. Promoting AI applications has become one of the focal points of many companies' strategies [12]. The notable changes made by AI have inspired recent studies to examine the impacts and consequences of the technology and investigate AI's performance implications. However, this objective requires an in-depth understanding of the factors affecting the acceptance of AI applications by potential users in different manufacturing and service fields.

AI in Health Care

Previous studies highlight the importance of AI in health care, especially in medical informatics [13]. AI can improve patient care, diagnosis, and interpretation of medical data [14]. Houssami et al [15] showed that AI applications used for breast cancer screening reduced human detection errors; however, some of the interrelated ethical and societal trust factors, as well as reliance on AI, are yet to be developed. Prior research has shown that AI clinical applications exhibit the same or sometimes even better performance than their human counterparts or specialists in detecting Alzheimer disease using natural language processing techniques [16], and for detecting skin cancer [17] and heart arrhythmia [18] using deep neural networks. AI applications for health care recommendations may differ from those in other sectors, mainly because of the highly sensitive nature of health information and high levels of consumer vulnerability to possible medical errors.

In the context of AI in health care, there can be three possible patient encounters for care delivery. First, the patient can follow the traditional health care delivery model and visit a physician in person. This option is the most prevalent health care delivery process that focuses on the physician-human interaction and can be referred to as a "human-human interaction" or "traditional in-person visit." Second, a patient can choose a collaborative intelligence [19] scenario where the physicians collaborate with an AI application to arrive at conclusions and make medical decisions. In other words, the physician's thinking, analysis, and performance are augmented by using AI applications to interact with patients, but the ultimate responsibility of patient care and recommending treatment options and care planning still rest with the physician. Third, AI clinical applications substitute the physician, and a patient only encounters the AI clinical application unaided by human intelligence.

Acceptance of AI in Clinical Applications

In April 2018, the Food and Drug Administration authorized the first AI application to diagnose diabetic retinopathy without a physician's help in the United States [20]. An increasing number of health care service companies have invested in AI applications in mobile health devices or health apps to improve patient safety, increase practice quality, enhance patient care management, and decrease health care costs. However, previous studies suggest that not all individuals are willing to accept AI clinical applications [20]. Successful implementation of AI applications requires a careful examination of users' attitudes and perceptions about AI [21]. Thus, investing in AI applications without recognizing potential users' beliefs and willingness to use them may waste resources and even result in customer loss. This is especially true in the health care sector, where patient engagement is considered one of the most critical determinants of health care quality. If individuals do not view interacting with AI clinical applications as useful, they may demand interactions with physicians, and in turn, the AI applications may remain unused. Therefore, understanding the decision drivers and barriers that lead to acceptance or refusal of AI clinical applications in health care delivery is fundamental for health care providers and hospitals that plan to introduce or increase AI presence during health care delivery.

Several studies have investigated the attitude of specific sample populations toward AI in health care. For example, Dos Santos and colleagues [22] evaluated medical students' attitudes toward AI and found that the majority agreed that AI could improve medicine as a whole. In a similar study involving medical

students in Canada [23], the majority of the sample (67%) agreed that AI would reduce the demand for medicine and 47% were anxious about physicians' future in association with AI. In a survey conducted among members of the European Society of Radiology [24], more than half of the respondents strongly expressed that they were not ready to accept AI-only generated reports. Similarly, in a study investigating perceptions toward health care robots [25], residents in a retirement village exhibited a more positive attitude toward the robot compared to the responses of their staff and relatives. Recently, Patel and colleagues [26] compared the diagnosis results of pneumonia on chest radiographs among human experts alone, collaborative intelligence, and two state-of-the-art AI deep learning models. They demonstrated that both the collaborative and AI models have superior performance when compared with human experts alone. They also found that the combination of collaborative and AI models outperforms each of these methods alone.

Research Gaps

Based on previous studies, health care professionals still express fundamental concerns about implementing AI clinical applications in care services [27-30]. These concerns and risks also directly affect patients' perceptions (as potential users and beneficiaries) and make them withdraw from using AI clinical applications [20]. The majority of studies investigating attitudes toward AI involve national samples [31], medical students [22], radiology experts [23], and physicians [27], and we did not find a study that examined the perceptions of patients with different health conditions (chronic and acute diseases) toward AI clinical applications. Exploring the perspectives of individuals with different diseases can allow researchers to effectively recognize the source of risks and concerns associated with AI clinical applications. Since the risks may vary by type of illness, this perspective can help health care providers better understand how to address the concerns of various patients.

Thus, researchers need to understand the current challenges related to AI technologies more efficiently and analyze the urgent needs of health systems to design AI applications to address them. Even with physicians and other health care stakeholders accepting and assimilating AI to varying degrees, it is crucial to understand patients' perspectives toward these different scenarios. Nevertheless, little is known about the risk beliefs associated with using AI clinical applications for diagnosis and treatments from the general public's perspective. This stream of literature encouraged us to examine people's perceptions and attitudes toward different types of health care delivery processes in this study.

Currently, the issues related to AI clinical applications in health care are still within the realm of research. However, it is widely believed that these systems will fundamentally change medical practice in the near future [32]. Historically, the medical sector does not integrate technology as quickly as other industries [33]. Moreover, integrating AI into the current medical workflow could be very challenging without the involvement, cooperation, and endorsement of stakeholders (such as health care professionals and patients) and a robust legislative and regulatory framework. The main objective of this study was to examine how potential users perceive the benefits, risks, and use of AI clinical applications for their health care purposes, and how their perceptions may be different if faced with the three health care service encounter scenarios. The benefit perceptions and risk beliefs of prospective users may affect their future adoption of AI applications. Patients may not decide what tools health care professionals should use in their practice, but they can definitely highlight possible concerns, challenges, and barriers that may refrain them from supporting and using the tools implemented and promoted by clinicians.

Literature Review

Overview

In this section, consistent with the research objectives, three topics are explained. First, the type of illness and the reactions of people with different illnesses (acute or chronic) are described. Second, possible risks and concerns associated with the use of AI clinical applications are highlighted. Third, potential benefits that users may perceive from using AI in health care are illuminated. The interrelationships among these three topics can provide further research background on how people with different health conditions may react to AI applications used for health care purposes to place our research objectives and experimental design in context.

Type of Illness (Acute or Chronic)

A patient may experience two general types of health conditions: acute diseases and chronic diseases. Following the medical literature, acute conditions are defined as diseases that develop suddenly, are severe and sudden in onset (the initial phase of a disease or condition in which symptoms first become apparent), last a short time (often only a few days or weeks), and can be cured [34]. In contrast, a chronic disease is described as a human health condition or disease that is persistent or otherwise long-lasting in its effects or a disease that develops over time [35]. Thus, chronic diseases refer to long-term health conditions that last more than 1 year [36], whereas acute diseases refer to health conditions that are sudden, short-term, and require medical attention. Examples of acute diseases include the common cold, flu, and infections, whereas examples of chronic diseases include Alzheimer disease, arthritis, diabetes, and depression [36]. Given the contrasting nature of chronic and acute conditions, it is logical to argue that patients with different diseases will vary in their perceptions of AI in health care delivery. Previous research indicates that individuals tend to trust an algorithm or an AI system in low-risk conditions [37]. Based on that finding, we can expect those with acute short-term conditions but in severe pain to opt for an AI clinical encounter. For example, Wu and colleagues [38] explored older adults' perceptions of mild cognitive impairment toward assistive AI and found a generally positive belief that these AI applications can be useful for the aging population.

Perceived Concerns and Risks

Perceived Communication Barriers

Conventionally, the health care delivery process usually occurs in a hospital or a physician's clinic and involves direct physician-patient interaction, which can be described as paternalistic in nature. In other words, with medical expertise,

the physician leads the patient toward shared medical decision-making, resulting in outcomes such as prescription and treatment plans centered on both evidence-based medicine and moral competency in terms of showing empathy and compassion. Empathy and compassion are increasingly being viewed as the foundations of active patient engagement and patient-centered care [39,40]. Research highlights that patients mainly seek and trust health care providers who are competent and compassionate with good interpersonal skills [41-44]. Researchers also reveal that empathetic and compassionate physician-patient interaction can improve patient satisfaction and greater clinical adherence [45]. Conversely, AI applications in service delivery (such as health care) may cause noteworthy communication barriers between customers and AI applications [46]. Reliance on AI clinical applications may reduce physicians' and patients' interactions and conversations [47]. Consumers may refuse to use AI applications because they need human social interaction during service encounters [10]. AI applications are powered with higher-level technical and evidence-based medicine but may not be expected to exhibit human-like empathy, which, in turn, may discourage patients from choosing the AI applications for the health care delivery process. Although the idea of building empathetic machines is well pursued in AI, patients' perceptions toward the clinical encounters involving AI as substituting or augmenting technology versus traditional in-person visits warrant further investigation.

Perceived Transparency of Regulatory Standards

Physicians obtain their licensure after many rigorous training years in medicine and their specialty. This licensure is considered a regulating mechanism put forward by the government to ensure physicians' quality and, ultimately, the quality of health care services. The licensure further allows a patient to choose a reliable doctor responsible for intentional errors or unintentional wrong-doings. However, in the AI context, regulatory authorities are yet to formalize standards to evaluate and maintain AI's safety and impact in many countries [48]. Thus, people may become concerned if an appropriate regulatory and accreditation system regarding AI clinical applications is not yet in place. In the ever-changing AI and machine-learning landscape, more effective, efficient, and powerful algorithms are being developed on an everyday basis to power these AI health care applications. Often, the technical aspects of modern AI algorithms such as artificial neural networks (ANNs) remain a black box to society at large [49]. This is because after an ANN is trained with a data set, the quest to understand the algorithm's decision-making process becomes essentially impossible. This perception of the "unknown" could potentially affect a patient's preference for the clinical encounter. Furthermore, the lack of transparency in the regulatory standards that can be understood, critiqued, and reviewed [50] for AI applications by a larger community may discourage patients from choosing an AI encounter. The new IEEE (Institute of Electrical and Electronics Engineers) standard P7001 currently under development, "Transparency in autonomous systems," has a set of measurable and testable levels of transparency that could be used to evaluate AI or autonomous systems for their level of compliance [51].

Perceived Liability Issues

A physician is usually held responsible for the consequences of their actions in a health care setting. With AI increasingly finding its way into health care research and practice, it becomes imperative to examine an AI system's liability issues. Previous studies in public health demonstrate legal concerns about who will account for AI-based decisions when errors occur using AI applications [52]. Usually, the stakeholders in a medical encounter involving AI can be the developers, data feeders, health care organizations that adopted AI, or the health care provider that used the AI [53]. Noting that an AI application in itself cannot be held liable for any misdiagnosis or medical recommendations that turn out to be disastrous for a patient, the lack of standard consensus or regulations on who can be held liable may discourage patients from choosing an AI application. As AI clinical applications make autonomous decisions, the accountability question becomes very hard to answer. For instance, it will create a risky situation for both clinicians and patients when it is still unclear who becomes responsible if AI clinical applications offer wrong health care recommendations [54]. There is also no precise regulation regarding who is held liable when a physician follows the medical recommendations provided by AI and when a physician decides to override the recommendations [55].

Perceived Trust in AI Mechanisms, Collaborative Intelligence, and Physicians

Maintaining substantial trust between the public, health professionals, and health systems can create effective health care. Trust can be defined as trust in clinicians and the clinical tools they use (such as AI clinical applications) [48]. In the information systems (IS) literature, Vance and colleagues [56] call for additional research on trust in information technology artifacts such as AI systems. Gaining the general public's trust in the use of AI in health care is considered an important challenge to the successful implementation of AI in medical practices [57]. Sun and Medaglia [58] reported that, in general, individuals are likely to exhibit a lack of trust in the features of AI systems. For instance, people may not trust AI's predictive power and diagnostic ability for treatment purposes. Another study indicated that the autonomy of AI systems affects the users' perception of trustworthiness [59]. Moreover, in a survey conducted by Longoni and Bonezzi [60] to understand customer perceptions about AI in medicine, only 26% of the sample signed up for an AI diagnosis compared to 40% who signed up for a health care provider diagnosis. In the same study, the authors found that the majority preferred a human provider over AI, even when it meant a higher risk of misdiagnosis. They also indicated that patients are more willing to choose AI if the ultimate treatment decision rests with the physician and not only the AI. These results highlight that individuals have a higher level of distrust toward AI in medicine than toward a human provider. Trust in AI clinical applications is a significant factor affecting adoption decisions [61]. Longoni et al [60] suggested that a physician's confirmation of the AI results (an example of AI as augmenting technology) could encourage patients to be more receptive to AI in their care.

Perceived Performance Risks (Possible Errors)

AI-related studies consider the safety and quality of autonomous operations as essential factors affecting the use of AI applications [62]. According to Mitchell [63], AI applications are still vulnerable in many areas such as hacker attacks. Hackers can change text files or images, which may not have a human cognitive effect but could cause potentially catastrophic errors. Since the AI program may not understand the input and outputs, they are susceptible to unexpected errors and untraceable attacks. Performance risks may be serious in the context that directly deals with people's lives (such as health care). Medical errors generated by AI could endanger patient safety and result in death or injuries, which are mostly not reversible. Thus, users may be concerned that the mechanisms used by AI clinical applications could lead to incorrect diagnoses or wrong treatments. Reddy et al [54] indicated that incomplete and nonrepresentative data sets in AI models can produce inaccurate predictions and medical errors. Thus, it could be expected that individuals may consider that possible functional errors resulting from using AI applications could lead to more risks.

Perceived Social Biases

Studies in other contexts have shown that AI models overestimate crime risk among members of a specific racial group [64]. In the health care context, biased AI models may overestimate or underestimate health risks in specific patient populations. For instance, AI applications may engage in stereotyping and exhibit gender or racial bias. Bias in AI models may also occur when data sets are not representative of the target population or when AI systems use incomplete and inaccurate data for decision-making [47]. Societal discrimination (such as poor access to health care) and small samples (such as minority groups) can lead to unrepresentative data and AI bias [48]. Edwards [65] argued that AI systems' current architecture needs a more sophisticated structure to understand human moral values. If the AI algorithm is not transparent, it may exhibit some discrimination levels, even though humans are not involved in decision-making [66]. The main purpose of AI is to create an algorithm that functions autonomously to find the best possible solutions to questions [67]. However, researchers argue that predictive programs can be inevitably biased due to an overrepresentation of the social minorities in the pattern recognition process [68]. Some studies support this argument by showing that AI algorithms may be coded in a biased manner, which can produce racist decisions [69]. Therefore, if people are concerned that AI applications could lead to morally flawed health care practices by overestimating or underestimating health risks in a certain patient population, they will be more likely to perceive greater risks associated with AI.

Perceived Privacy Concerns

Health-related data are often viewed as constituting the most sensitive information about a person [47]. In health care services, respecting a person's privacy is an essential ethical principle because patient privacy is associated with well-being and personal identity [70]. Thus, patients' confidentiality should be respected by health care providers by protecting their health records, preventing secondary use of data, and developing a robust system to obtain informed consent from them for health care purposes [71]. If patients' privacy needs are not met, patients will be affected by psychological and reputational harm [72]. Data breaches would increase risk beliefs associated with AI models designed to share personal health information. There is a concern that anonymized data can be reidentified through AI processes, and this anxiety may exacerbate privacy invasion and data breach risks [48]. AI applications in public health require large data sets. Thus, collecting, storing, and sharing medical data raise ethical questions about safety, governance, and privacy [73]. Privacy is one of the most critical concerns associated with using AI applications because users' data (eg, habits, preferences, and health records) are likely to be stored and shared across the AI network [66]. The method of data collection for AI may increase risks as AI systems need huge data sets, and patients are concerned that their personal information will be collected without their knowledge [47].

Perceived Benefits

AI can be used in health care for risk prediction and recommendation generation. Big data and AI significantly improve patient health-based diagnosis and predictive capability [74]. Recent studies highlight new AI application opportunities within medical diagnosis and pathology, where medical tasks can be performed in an automated manner with higher speed and accuracy [75]. AI can improve health care delivery such as diagnostics, prognosis, and patient management [48]. For instance, AI has been shown to be capable of diagnosing skin cancer more efficiently than dermatologists [76]. Sohn and Kwon [77] demonstrated that hedonic aspects such as enjoyment and curiosity about AI technology are stronger in predicting the behavioral intention to use AI products than utilitarian aspects (eg, usefulness). This point does not hold in health care since AI applications are mainly used in health care for utilitarian aspects such as patient-specific diagnosis, treatment decision-making, and population risk prediction analysis [78]. Thus, with regard to benefit perceptions, in this study, we only focus on utilitarian aspects, not other motivational factors. Sun and Medaglia [58] proposed the lack of sufficient knowledge of the AI technologies' values and advantages as potential barriers to adopting AI applications. Individuals will endorse and use AI clinical applications if they believe that AI will bring essential benefits to their health care delivery. Thus, we can expect that the higher the perceived benefits from AI clinical applications, the higher the individuals' intention to use them in the future.

Research Objectives

Most AI-related studies use various acceptance models (eg, technology acceptance model [TAM] and unified theory of acceptance and use of technology) to examine AI acceptance by empirically testing the effects of the ease of use, usefulness, and social norms on the intention to use AI applications [10,77]. For example, Xu and Wang [79] used the TAM to examine the adoption of AI robot lawyer technology for the legal industry. Another example is use of a TAM-based tool to measure AI-based assessment acceptance among students [80]. However, to the best of our knowledge, no experimental research has explored the differences in patients' perceptions (with different types of illnesses) in relation to utilizing AI clinical applications

without physician interactions, AI clinical applications with physician interactions, and traditional in-person visits. We hypothesized that the interactions between the type of health care service encounters and illness type may significantly change patients' perceived risks, benefits, and overall attitudes toward health care delivery. The main objectives of this study were to: (1) examine the difference between the perceptions of patients with chronic diseases about utilizing AI clinical applications with and without physician interactions, and visiting physicians in person for diagnosis and treatment recommendation purposes; (2) investigate the difference between the perceptions of patients with acute diseases about utilizing AI clinical applications with and without physician interactions, and visiting physicians in person for diagnosis and treatment recommendation purposes; and (3) explore which service encounter is preferable for people with chronic diseases and that desired by people with acute diseases.

In this study, we focused on AI clinical applications for health care purposes. AI embedded in mobile health devices or health apps could help patients monitor their health status, check their health care information, and manage their chronic illnesses. These AI clinical applications use algorithms to learn from the past by analyzing the medical histories of patients with the same health conditions; recognizing patterns in clinical data; predicting possible health issues; and suggesting some treatment choices, diagnostic options, prescription advice, and care planning. These applications could reduce frequent patient-physician encounters and avoid unnecessary hospitalizations.

Study Significance

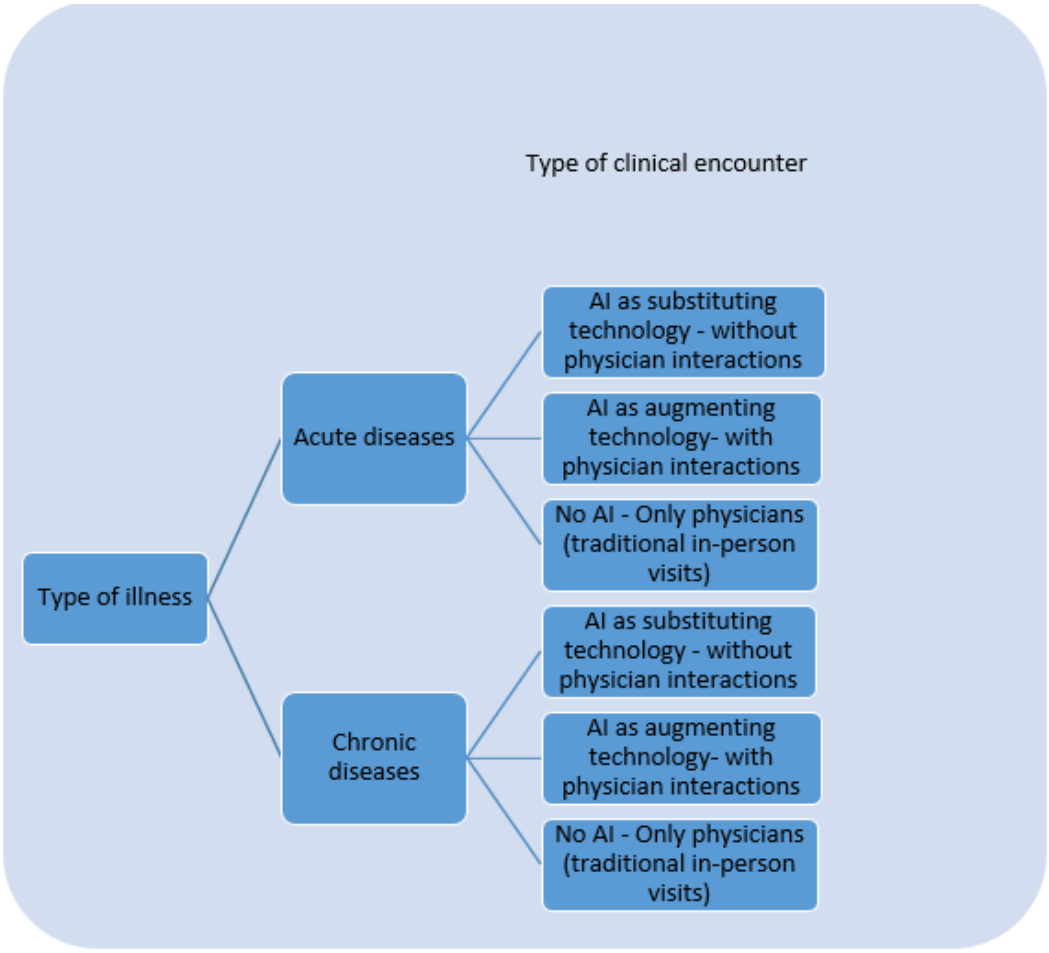
This research offers significant and timely insight into human-computer interaction by examining AI applications in health care. This study's findings will provide researchers and managers with critical insights into the determinants of individuals' intention to use AI applications in health care delivery. The results imply that incompatibility with instrumental, technical, ethical, or regulatory values can be a reason for rejecting AI applications in health care. Multidimensional concerns associated with AI clinical applications may also be viewed as a cause of technostress, which occurs when an individual is unable to adapt to using technology [81]. In the future, it will be the patient's or customer's right to choose AI-driven recommendations over human care or vice versa. Nevertheless, we propose that AI application developers and programmers devise practical strategies to anticipate possible concerns and minimize risk beliefs to encourage individuals to use AI technology for health care purposes. Our results highlight that patients may have various reactions to AI (as substituting or augmenting) technology. Thus, different strategies and policies may need to successfully implement AI as a substituting or as an augmenting technology to address potential concerns and risks.

Methods

Study Design

To understand patients' perceptions of AI applications in health care, we designed a 2×3 experiment that crossed a type of health condition (ie, acute or chronic) with three different types of clinical encounters (ie, AI clinical applications as substituting technology, AI clinical applications as augmenting technology, and no AI [traditional in-person visit]). In each scenario, we included two essential pieces of information: (1) the type of health condition and (2) the type of clinical encounter. We propose that the interactions between these two elements could lead to a comprehensive evaluation of how patients with different health conditions would perceive AI clinical applications in health care. It should be mentioned that participants of this study were actually suffering from either a chronic or acute disease. First, individuals entered their signs and symptoms to highlight what diseases they are suffering from. A filtering question was included at the beginning of the survey to categorize patients into the chronic or acute group. Each group was then randomly assigned to a hypothetical clinical encounter. For instance, an individual with an actual chronic illness was given a hypothetical situation in which they could use AI clinical applications under the physician's control. Figure 1 illustrates the six scenarios resulting from two types of diseases and three types of clinical encounters.

Figure 1. Six scenarios assigned to the study participants. AI: artificial intelligence.



In this study, we considered health conditions as either acute or chronic conditions. Acute diseases come on rapidly and are accompanied by distinct symptoms requiring urgent or short-term care, followed by improvement after treatment. For example, a broken bone that might result from a fall must be treated by a doctor and will heal in time. In some cases, an acute illness such as the common cold will simply go away on its own. Most people with acute illnesses will recover quickly. By contrast, chronic conditions develop slowly and may worsen over an extended period, from months to years. Chronic conditions are slower to develop, may progress over time, and may have many warning signs or no signs at all. Some examples of common chronic conditions are arthritis, diabetes, chronic heart disease, depression, high blood pressure, high cholesterol, and chronic kidney disease. Unlike acute conditions, chronic health conditions cannot be cured, only controlled and managed.

Regarding the second factor (clinical encounters), we focused on three categories: AI clinical applications as substituting technology, AI clinical applications as augmenting technology, and no AI (traditional in-person visit). In the AI as substituting technology scenarios (Scenarios 1-1 and 1-2), we defined a setting where patients directly use AI clinical applications for health care purposes. In these scenarios, we described a situation in which individuals can use an AI clinical application when they are suffering from a disease. The steps of using AI applications were clearly explained to respondents. For instance, when feeling sick, they can directly enter their signs, symptoms, and critical health complaints into the AI clinical application. Their health information will be recorded in a large database. The AI system then analyzes their health data, compares them to the learned patterns (eg, the list of diseases and medicines), and draws some clinical conclusions. Finally, based on the pattern found, the AI creates a report including some diagnostic options, some treatment choices, prescription advice (eg, dose, frequency and names of medications they need to take), care planning (eg, resting at home, taking suggested medicines for a specific period, or visiting a professional immediately). In summary, we highlighted that AI clinical applications could analyze clinical data and make medical decisions for patients without direct physician interactions. Therefore, we can consider this scenario using AI clinical applications without physician interactions to treat acute diseases or control chronic diseases.

In the AI as augmenting technology (with physician interaction) scenarios (Scenarios 2-1 and 2-2), we defined a setting in which patients have the option of using an AI clinical application that is monitored and controlled by their physician. In these scenarios, the physician will check the results and recommendations generated by the AI clinical application; discard some of the recommendations based on their experience and expertise; and make the final decision about the treatment choices, prescription options, and care planning. In summary, we emphasized that AI clinical applications can analyze clinical data and help physicians make medical decisions for patients. Thus, in this case, AI clinical applications are used with

physicians' direct supervision, and we can consider this scenario as AI-physician interactions (interactions and collaborations between physicians and AI).

In the no AI (only physician) scenarios (Scenarios 3-1 and 3-2), a patient has the option of visiting their physician. First, the physician asks for the patient's signs, symptoms, and critical health complaints through a conventional in-person visit. The physician analyzes collected health information. Then, based on the physician's knowledge, expertise, and experience, they draw some clinical conclusions. In summary, we highlighted that through a face-to-face patient-physician encounter, the physician can analyze clinical data and make final medical decisions. Thus, we can consider this scenario as an in-person visit with physicians. Table 1 displays the 2×3 experimental design of the six scenarios used in this study and the treatment group/scenario names.

Table 1. Research design overview.

Type of illness	AI[a] as substituting technology without physician interaction	AI as augmenting technology with physician interaction	Traditional in-person visit
Acute: temporary, short-term diseases	Scenario 1-1: Acute-AI-only	Scenario 2-1: Acute-AI-physician	Scenario 3-1: Acute-physician-only
Chronic: long-lasting diseases	Scenario 1-2: Chronic-AI-only	Scenario 2-2: Chronic-AI-physician	Scenario 3-2: Chronic-physician-only

[a]AI: artificial intelligence.

Each experiment included three sections. First, the scenario was described, which detailed each experiment's purpose (eg, in Scenario 1-2, the objective was defined as using AI clinical applications to control and monitor a chronic disease). Second, a set of questions about the nine outcome variables was provided to evaluate the respondents' perceptions based on the given scenario. For example, subjects were asked to reflect on possible trust issues with AI applications (in Scenarios 1-1 and 1-2), with collaborative intelligence (in Scenarios 2-1 and 2-2), and with physician interactions (in Scenarios 3-1 and 3-2). Finally, we asked our subjects to provide some demographic information.

Question Development

The main aim of this study was to evaluate individuals' perceptions of the health care service options described by six scenarios. We used the following variables to measure patients' perceptions: perceived performance risks, perceived communication barriers, perceived social biases, perceived privacy concerns, perceived trust, perceived transparency of regulatory standards, perceived liability issues, perceived benefits, and intention to use. We primarily included these variables to highlight the main barriers and facilitators of using AI clinical applications indicated by previous research [28]. Some variables such as perceived communication barriers and liability issues may have shared effects on the physician-patient interaction. However, in this study, we only focused on the impact perceived by patients. This study drew on the existing literature to measure the nine outcome variables used in the experiments, and minor changes were made to the questions to fit the given context (scenario). This study adapted items to measure outcome variables from existing scales developed by studies mainly conducted in the AI and medical fields. The descriptions of scenarios and final measure items used in this study are listed in Multimedia Appendix 1. Table 2 shows the definitions of all outcome variables used in this study.

Table 2. Operationalization of outcome variables.

Outcome variables	Variable definition	Reference
Perceived performance risks	The degree to which an individual believes that the clinical encounter (which is explained in the scenario) will exhibit pervasive uncertainties	Marakanon and Panjakajornsak [82]
Perceived communication barriers	The degree to which an individual feels that the clinical encounter (which is explained in the scenario) may reduce human aspects of relations in the treatment process	Lu et al [46]
Perceived social biases	The degree to which a person believes that a clinical encounter (which is explained in the scenario) may lead to societal discrimination to a certain patient group (eg, minority groups)	Reddy et al [48]
Perceived privacy concerns	The extent to which individuals are concerned about how the clinical encounter (which is explained in the scenario) will collect, access, use, and protect their personal information	Zhang et al [83]
Perceived trust	The degree to which an individual believes that the clinical encounter (which is explained in the scenario) is trustworthy	Luxton [55]
Perceived transparency of regulatory standards	The extent to which an individual believes that regulatory standards and guidelines to assess the safety of the clinical encounter (which is explained in the scenario) are yet to be formalized	Cath [70]
Perceived liability issues	The extent to which an individual is concerned about the liability and responsibility of using the clinical encounter (which is explained in the scenario)	Laï et al [20]
Perceived benefits	The extent to which an individual believes that the clinical encounter (which is explained in the scenario) can improve diagnostics and care planning for patients	Lo et al [84]
Intention to use	The extent to which an individual is willing to use the proposed clinical encounter (which is explained in the scenario) for diagnostics and treatments	Turja et al [27]

Since this study's subjects were individuals, we took two steps to ensure that the definitions, given scenarios, and questions were illustrated to be understandable for the general public. First, once the initial scenarios and surveys were developed, we consulted three professionals in the AI domain and two physicians (who were familiar with AI clinical applications) to improve our study's content validity and finalize the definitions, scenarios, and questions used in each survey. Consistent with the experts' suggestions, we modified the terms used to describe AI clinical applications, AI-physician interaction, as well as in-person examination, and improved the scenarios and questions to ensure that they were sufficiently transparent and easy to understand for the public. Second, we performed a face validity evaluation with 14 students (2 doctoral students in computer science, 1 doctoral student in IS, 4 master's students in computer science, 5 master's students in IS, and 2 medical students) to ensure that the readability of the scenarios and wording of the questions were acceptable and consistent with the objectives of our study. Thus, we reworded some ambiguous terms and removed technical language and jargon to describe the scenarios and develop the surveys in an understandable manner. It should be mentioned that graduate students may have a higher reading level than an average person. However, they were asked to detect and flag technical expressions and ambiguous terms that might not be clear to an average person. Therefore, the graduate students used more scrutiny to focus on every detail to ensure the questions were sufficiently transparent for our potential sample.

Data Collection

This study was reviewed and approved by the Institutional Review Board of Florida International University, and the data collection was performed confidentially. Written informed consent was obtained from all participants. All methods used in this study were carried out in accordance with relevant guidelines and regulations.

We used a power analysis to identify the appropriate sample size per scenario. The results of the power analysis showed that for a range of medium (0.5) to high (0.8) effect size [85], with α=.05 and power of more than 0.8, the total minimum sample required is about 50 respondents per scenario. In this study, there were nine main outcome variables with 49 measures. Therefore, to reduce possible sampling errors, we initially collected a sample of 121 respondents per scenario to ensure an adequate sample size after data cleaning and matching respondents in different scenarios. Data were collected in May 2020 from Amazon's Mechanical Turk (MTurk) to obtain a representative group of subjects in the United States. MTurk is a survey tool used in previous research that is considered as an acceptable means to collect individual-level data from the general population of interest [86]. The surveys of six scenarios were posted to MTurk, and the respondents' location was limited to the United States. We enabled a microcode in the survey design to prevent respondents from taking each survey more than once. Following previous studies that used MTurk for data collection, a monetary reward (US $0.70) was given as the incentive for participation. The range of average completion time for the six experimental groups was between5 minutes, 17 seconds and 8 minutes, 49 seconds, which indicated acceptable responses in terms of the time spent on each survey by the participants.

Data Analysis

IBM SPSS Statistics V21.0 was used to analyze the data. Propensity score matching was used with a tolerance of 0.05 to match participants and avoid any demographic bias between scenarios. To find each outcome variable's total score, we calculated unweighted sum scores of items for each variable.

Analysis of variance (ANOVA) was then performed to examine the differences between the six proposed scenarios for each of the outcome variables: perceived performance risk, perceived social biases, perceived privacy concerns, perceived trust, perceived communication barriers, perceived concerns about the transparency of regulatory standards, perceived liability issues, perceived benefits, and intention to use. Prior to ANOVA, the Levene test was run to ensure the homogeneity of variance, as this is one of the fundamental assumptions of ANOVA. There was sufficient evidence to hold the assumption of homogeneity of variance for all outcome variables. The Scheffe posthoc test was used to identify which scenarios significantly differed from each other per outcome variable.

Results

After cleaning the data for biases and incomplete responses, there were a total of 634 completed surveys. After matching across scenarios, there were 105 participants in Acute-AI-only, 104 participants in Chronic-AI-only, 113 participants in Acute-AI-physician, 103 participants in Chronic-AI-physician, 105 participants in Acute-physician-only, and 104 participants in Chronic-physician-only. The detailed demographic information of the six scenarios is reported in Multimedia Appendix 2. In summary, 44% of participants were women; approximately 30% of participants were between 20 and 29 years old, 33% were between 30 and 39 years old, 18% were between 40 and 49 years old, and 17% were above 50 years of age. The majority of the participants were White (65%), followed by 17% Asian, 11% African American, and 5% Hispanic. Regarding the level of education, 6% were high school graduates, 11% completed some college, 8% held a 2-year degree, 48% had a bachelor's degree, and 23% had a master's degree. Regarding employment status, most of the participants were full-time employees (72%), followed by 14% part-time employees, 8% unemployed, 2% retired, and 4% students. Approximately 15% of participants in our study reported an annual household income of less than US $25,000, 26% reported an income between US $25,000 and US $49,999, 22% reported an income between US $50,000 and US $74,999, 18% reported an income between US $75,000 and US $99,999, and approximately 19% reported an income of more than US $100,000.

Across the six scenarios, there were no significance differences in terms of gender (χ^2_5=1.76, *P*=.88), age (χ^2_{25}=30.31, *P*=.21), race (χ^2_{25}=22.37, *P*=.62), level of education (χ^2_{30}=37.89, *P*=.15), employment (χ^2_{20}=16.20, *P*=.70), and annual household income (χ^2_{25}= 19.85, *P*=.76). Respondents were also asked to report their personal innovativeness on a Likert scale to ensure this factor would not introduce any bias into different scenarios. The ANOVA results across the six scenarios revealed no significant differences among respondents regarding the level of personal innovativeness (*P*=.19).

The items were adapted from previous studies with slight changes to fit them into this research context. All items were measured on a 5-point Likert-type scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree." Table 3 shows the number of items and Cronbach α values per outcome variable, which were all above .70 as the recommended threshold value [87], implying adequate reliability per outcome variable.

Table 3. Reliability of variables.

Outcome variables	Number of items	Cronbach α
Perceived performance risks	5	.92
Perceived social biases	4	.85
Perceived privacy concerns	6	.93
Perceived trust	5	.92
Perceived communication barriers	5	.92
Perceived transparency of regulatory standards	5	.92
Perceived liability issues	6	.93
Perceived benefits	7	.92
Intention to use	5	.92

The summary statistics (mean score, SD) per outcome variable are presented in Table 4. Some of the trends are evident from these results. For example, we can observe lower privacy concerns and liability issues for traditional in-person examinations than AI-based interactions.

Table 4. Summary statistics of outcome variables as a function of a 2 (type of illness) by 3 (type of encounter) design.

Outcome variable	AI[a] as substituting technology (without physician interaction), mean (SD)	AI as augmenting technology (with physician interaction), mean (SD)	Traditional in-person visit, mean (SD)	ANOVA[b]	
				F statistic (*df*=5, 628)	*P* value
Perceived performance risks				1.36	.24
Acute, short-term illness	16.3 (5.1)	16.6 (5.5)	15.4 (5.0)		
Chronic, long-lasting illness	16.2 (5.1)	17.1 (4.8)	15.9 (5.3)		
Marginal means[c]	16.3 (5.1)	16.8 (5.1)	15.6 (5.1)		
Perceived biases				0.86	.51
Acute, short-term illness	12.8 (3.7)	13.0 (4.4)	12.2 (3.8)		
Chronic, long-lasting illness	13.0 (3.9)	13.2 (3.8)	12.4 (4.3)		
Marginal means	12.9 (3.8)	13.1 (4.1)	12.3 (4.0)		
Perceived privacy concerns				3.35	.005
Acute, short-term illness	19.0 (6.7)	20.8 (6.1)	17.7 (6.1)		
Chronic, long-lasting illness	20.5 (5.8)	19.8 (6.1)	19.5 (7.0)		
Marginal means	19.8 (6.3)	20.3 (6.1)	18.6 (6.6)		
Perceived trust				6.27	<.001
Acute, short-term illness	16.5 (5.0)	17.3 (4.8)	18.6 (4.4)		
Chronic, long-lasting illness	15.4 (4.9)	16.8 (4.7)	18.2 (4.8)		
Marginal means	16.0 (5.0)	17.1 (4.8)	18.4 (4.6)		
Perceived communication barriers				9.24	<.001
Acute, short-term illness	17.4 (5.7)	18.1 (5.2)	14.7 (4.8)		
Chronic, long-lasting illness	17.1 (4.9)	17.5 (4.8)	14.6 (5.7)		
Marginal means	17.3 (5.3)	17.8 (5.0)	14.6 (5.3)		
Perceived transparency of regulatory standards				9.42	<.001
Acute, short-term illness	17.9 (5.0)	18.1 (5.1)	14.9 (4.9)		
Chronic, long-lasting illness	17.6 (4.7)	17.6 (5.0)	15.0 (5.5)		
Marginal means	17.8 (4.9)	17.8 (5.0)	14.9 (5.2)		
Perceived liability issues				6.27	<.001
Acute, short-term illness	21.3 (6.4)	22.1 (5.7)	18.5 (6.0)		
Chronic, long-lasting illness	20.8 (5.8)	20.6 (6.2)	18.3 (6.7)		
Marginal means	21.0 (6.1)	21.4 (6.0)	18.4 (6.4)		
Perceived benefits				3.28	.006
Acute, short-term illness	24.5 (6.2)	25.9 (5.5)	26.1 (5.9)		
Chronic, long-lasting illness	23.6 (6.5)	24.2 (6.6)	26.0 (6.2)		
Marginal means	24.1 (6.4)	25.1 (6.1)	26.1 (6.0)		
Intention to use				9.71	<.001
Acute, short-term illness	16.6 (4.8)	17.4 (4.9)	19.3 (4.6)		
Chronic, long-lasting illness	15.8 (5.3)	16.5 (5.2)	19.3 (4.6)		
Marginal means	16.2 (5.1)	17.0 (5.0)	19.3 (4.6)		

[a]AI: artificial intelligence.

[b]ANOVA: analysis of variance.

[C]Difference in the means of acute short-term illness and chronic long-lasting illness.

Significant differences (P<.05) between different groups were found for the following variables: perceived privacy concern, perceived trust, perceived communication barriers, perceived concerns about transparency in regulatory standards, perceived liability issues, perceived benefits, and intention to use. No significant difference was found between scenarios regarding perceived performance risk and perceived social biases.

Table 5 shows a summary of significant differences between scenarios from the Scheffe posthoc test. Patients suffering from an acute temporary short-term disease were significantly more concerned about the privacy of their health information when AI clinical applications with physician interaction was used compared to having a traditional in-person interaction with their physicians (P=.03). No significant differences were found in terms of perceived privacy concerns among patients with chronic conditions across scenarios.

Concerning trust, our results showed that patients with chronic illnesses found AI clinical applications to be less trustworthy compared to traditional diagnostic and treatment processes when they interact directly with the physicians (P=.004). The trust in physicians was also significantly higher for patients with acute conditions (P<.001).

Regarding perceived communication barriers, patients were significantly more concerned that AI clinical applications may reduce or eliminate the human aspect of relations between patients and professional care providers in comparison with face-to-face physician interactions for both acute (P=.01) and chronic (P<.001) health conditions. Similarly, when AI clinical applications are used in addition to physician interaction, there were still significantly greater concerns about lack of human relations than in face-to-face physician visits for both acute (P=.03) and chronic (P=.005) illnesses.

Further, the results showed that patients were significantly more concerned about the transparency of regulatory standards to assess AI algorithms and tools in comparison with the transparency of guidelines to monitor the performance of physicians' practices for both acute (P=.002) and chronic (P=.02) conditions. Similarly, when AI clinical applications are used in addition to physician interactions, patients were significantly more concerned about the transparency of guidelines for AI-physician interactions than traditional in-person physician visits for acute (P=.001) and chronic (P=.02) illnesses.

Regarding patients' concerns about liability issues, patients with acute illnesses were significantly more concerned when AI clinical applications are used under physicians' control. This may be because of the lack of clarity about who is responsible if appropriate AI-recommended treatment options are mistakenly dismissed or offer wrong recommendations compared to physician liability in traditional visits (P=.003). Interestingly, patients suffering from a chronic condition were significantly more concerned about liability issues using only AI clinical applications (P=.04) or AI tools with physician control (P=.001).

Lastly, patients with acute illnesses indicated significantly higher intentions to use in-person visits than only AI clinical applications (P=.01). By contrast, patients with chronic illnesses were significantly more willing to use in-person visits compared to only AI tools (P<.001) as well as AI clinical applications under physician control (P=.006). The detailed results, including nonsignificant differences, are included in Multimedia Appendix 3.

Table 5. Comparison of outcome variables between the six scenarios.

Scenarios compared	Mean difference (SE)	*P* value	95% CI
Perceived privacy concern			
Acute-AI-physician vs Acute-physician-only	3.07 (0.86)	.03	0.21 to 5.92
Perceived trust			
Chronic-AI-only vs Acute-physician-only	–3.22 (0.66)	<.001	–5.42 to –1.01
Perceived communication barriers			
Acute-AI-only vs Acute physician only	2.75 (0.72)	.01	0.36 to 5.14
Acute-AI-only vs Chronic-physician-only	2.86 (0.72)	.01	0.46 to 5.26
Chronic-AI-only vs Acute-physician-only	2.41 (0.72)	.05	0.01 to 4.81
Chronic-AI-only vs Chronic physician-only	2.52 (0.72)	.03	0.12 to 4.92
Acute-AI-physician vs Acute-physician-only	3.38 (0.70)	<.001	1.03 to 5.73
Acute-AI-physician vs Chronic-physician only	3.49 (0.71)	<.001	1.13 to 5.84
Chronic-AI-physician vs Acute-physician-only	2.87 (0.72)	.01	0.46 to 5.27
Chronic-AI-physician vs Chronic-physician-only	2.98 (0.72)	<.001	0.57 to 5.38
Perceived transparency of regulatory standards			
Acute-AI-only vs Acute-physician-only	3.04 (0.70)	<.001	0.71 to 5.36
Acute-AI-only vs Chronic-physician-only	2.91 (0.70)	<.001	0.57 to 5.24
Chronic-AI-only vs Acute-physician-only	2.77 (0.70)	.01	0.44 to 5.10
Chronic-AI-only vs Chronic-physician-only	2.64 (0.70)	.02	0.30 to 4.97
Acute-AI-physician vs Acute-physician-only	3.22 (0.68)	<.001	0.94 to 5.51
Acute-AI-physician vs Chronic-physician-only	3.09 (0.69)	<.001	0.80 to 5.38
Chronic-AI-physician vs Acute-physician-only	2.71 (0.70)	.01	0.37 to 5.04
Chronic-AI-physician vs Acute-physician-only	2.57 (0.70)	.02	0.23 to 4.91
Perceived liability issues			
Acute-AI-only vs Chronic-physician-only	2.92 (0.85)	.04	0.09 to 5.75
Acute-AI-physician vs Acute-physician-only	3.53 (0.83)	<.001	0.75 to 6.30
Acute-AI-physician vs Chronic-physician-only	3.73 (0.83)	<.001	0.94 to 6.51
Intention to use			
Acute-AI-only vs Acute-physician-only	–2.75 (0.68)	.01	–5.02 to –0.49
Acute-AI-only vs Chronic-physician-only	–2.73 (0.68)	.01	–5.00 to –0.46
Chronic-AI-only vs Acute-physician-only	–3.52 (0.68)	<.001	–5.79 to –1.25
Chronic-AI-only vs Chronic-physician-only	–3.49 (0.68)	<.001	–5.76 to –1.22
Chronic-AI-physician vs Acute-physician-only	–2.79 (0.68)	.01	–5.06 to –0.52
Chronic-AI-physician vs Chronic-physician-only	–2.76 (0.68)	.01	–5.04 to –0.48

Discussion

Principal Findings

Given the promising opportunities created by AI technology (such as better diagnostic and decision support), the main question is when AI applications will become part of routine clinical practice [88]. AI embedded in smart devices democratizes health care by bringing AI clinical applications into patients' homes [47]. Nevertheless, some concerns related to the use of AI need to be addressed. As previous studies introduced several concerns and challenges with AI [66], this study's main focus was to analyze the perceptions of people with different health conditions about the use of AI clinical applications as an alternative for diagnostics and treatment purposes. This study required participants who are actually suffering from an acute or chronic disease to consider hypothetical situations (ie, using AI clinical applications). If the study had recruited participants who are current users of AI clinical applications for health care purposes, the results could have been different. Current users of AI applications in health care settings may have more accurate perceptions about AI, and

the findings could be more practical. In the following subsections, we propose our theoretical contributions and practical implications related to each outcome variable.

Perceived Communication Barriers

The results showed that people with both acute and chronic health conditions may believe that both AI applications and collective intelligence can lead to communication barriers. This point is in line with previous studies highlighting that the use of AI applications in service delivery (such as health care) may cause noteworthy communication barriers between customers and service providers [46]. Reliance on AI clinical applications may reduce physicians' and patients' interactions and conversations [47]. Consumers may refuse to use AI applications because they need human social interaction during service encounters [10]. AI technology fundamentally changes traditional physician-patient communications. Thus, individuals may worry as they may lose face-to-face cues and personal interactions with physicians. AI creates challenges to patient-clinician interactions, as clinicians need to learn how to interact with the AI system for health care delivery and patients are required to reduce their fear of technology [89]. As AI continues to proliferate, users still encounter some challenges concerning effective use, such as how the partnership between AI systems and humans could be synergic [2]. A previous study proposed that more sophisticated technologies should be integrated into current AI clinical applications to improve human-computer interactions and streamline the information flow between two parties [66]. Therefore, the nature of AI clinical applications (even coupled with physician controls) may reduce conversation between physicians and patients, resulting in the emergence of more risk beliefs.

Suppose individuals are concerned that AI applications may reduce human aspects of relations in medical contexts. In that case, they may lose face-to-face cues and personal interactions with physicians and find themselves in a more passive position for making health-related decisions. This finding is consistent with a study in the chatbot context (within the area of AI systems), which indicated that users have stronger feelings of copresence and closeness when the chatbot uses social cues [90]. In the context of robot care, a study showed that when robots are used in rehabilitation, they are viewed by patients as reducing human contact [91]. Developers need to add more interactive and entertaining social cues to AI clinical applications to address the possible communication barriers between users and AI. For instance, AI-driven recommendations and assistance can be appealing if the application holds promise of allowing users more time to interact with it to establish empathy.

Perceived Privacy Concerns

In the case of suffering from an acute illness, people may perceive more serious privacy concerns if they can use AI clinical applications that are under physician control. Thus, they may prefer to use face-to-face interactions with physicians to reduce their privacy concerns. Deeper privacy concerns may have roots in two common perceptions. The first is the belief that anonymized data can be reidentified through AI models, and in turn, could increase the likelihood of privacy invasion and data breach [48]. The second is that AI systems need massive data sets; thus, patients are concerned that their health information may be collected or shared without permission for purposes other than treatment [47].

Perceived Trust

The findings imply that individuals with chronic conditions may not trust AI clinical applications if no physician interactions are included in health care delivery. According to previous studies, the nature of AI models (such as deep learning) may increase a lack of transparency related to AI systems and threaten patient trust, resulting in higher risk beliefs [57]. When patients cannot understand the inside workings of AI applications (such as decision-making models), they may exhibit lower trust in their functions and how they generate treatment solutions and recommendations. Thus, people with chronic diseases may be more willing to trust direct patient-physician interactions to control and manage their symptoms.

Perceived Accountability Issues

Accountability and liability are the other major concerns related to the use of AI. In this study, patients with acute conditions were more likely to be concerned about liability issues in the scenarios of both purely AI clinical applications and AI-physician interactions. Acute diseases are often accompanied by distinct symptoms that require urgent or short-term care. Thus, patients with severe and sudden signs and symptoms may seek quick care planning, accurate diagnosis, and reliable treatment options to cure their health problems promptly. In this situation, patients will become more nervous if they do not know who is held responsible for possible medication errors (such as wrong drug selection, wrong dose, or wrong quantity). This finding is consistent with previous studies in public health demonstrating the legal concerns surrounding who can be held accountable for AI-based decisions when errors occur using AI systems [52]. In general, society is yet to fully grasp many of the accountability and responsibility considerations associated with AI and big data [92]. Accountability involves several stakeholders such as AI developers, government agencies, health care institutions, health care professionals, and patient communities. Nevertheless, it is still not clear how the regulatory concerns around responsibility and accountability of using solutions made by AI systems can be dealt with formally [66]. Liability complexity becomes higher since it is not transparent to what extent AI systems can guide and control clinical practices [93]. Responsibility concerns are not only limited to the incidents in which AI may generate errors. Another aspect of liability risk is when the right and appropriate treatment options recommended by AI are mistakenly dismissed [55]. Thus, the higher the perceived liability issues, the greater the risk beliefs associated with AI. Regulatory agencies and health care organizations require clear policies to identify each stakeholder's responsibility (eg, patients, physicians, hospitals, and AI developers) when AI clinical applications are widely offered.

Perceived Transparency of Regulatory Standards

Regarding the regulatory risks associated with the transparency of standards, patients with either acute or chronic conditions

were concerned with using purely AI-based services as well as AI applications under physicians' direct supervision. This is in line with previous studies, which highlight that regulatory concerns are critical challenges to the use of AI in health care as the policies and guidelines for AI applications are not yet transparent [70]. The existing literature indicates that regulatory agencies require agreement on a set of standards that medical AI rollout must be rated against, such as determining the reliability of auditing the decisions made by autonomous AI clinical applications [66]. Due to the intelligence nature of AI systems, regulatory agencies should establish new requirements, official policy, and safety guidelines regarding AI rollout in health care [20]. For example, there is a legal need to evaluate the decision made by AI systems in case of litigation. AI applications operate based on autolearn models, which improve their performance over time [94]. This inner mechanism differentiates AI applications from other health care tools and gives rise to new regulatory concerns that may not be the case in different domains. Generally, algorithms that change continuously with features that are not limited to the original accepted clinical trials may need a new range of policies and guidelines [47]. Regulatory authorities are yet to formalize standards to evaluate and maintain AI's safety and impact in many countries [48]. Thus, people may become concerned if an appropriate regulatory and accreditation system regarding AI clinical applications is not yet in place.

The lack of clear guidelines to monitor the performance of AI applications in the medical context can lead to higher risk beliefs associated with AI. Hence, if health care organizations cannot reduce regulatory concerns, many individuals may refuse to use AI clinical applications and request traditional interactions with physicians. Even if hospitals decide to use AI applications as supportive services under health care professionals' supervision, the regulatory concerns should be mitigated prior to implementing AI systems. Regulatory agencies should establish normative standards and evaluation guidelines for implementing and using AI in health care in cooperation with health care institutions. The policies should clarify how AI clinical applications will be designed and developed in health care to comply with the accepted ethical principles (such as fairness and health equity). Regular audits and ongoing monitoring and reporting systems can be used to continuously evaluate the safety, quality, transparency, and ethical factors associated with services delivered through AI clinical applications.

Perceived Performance Risks

There were no significant differences found across the scenarios regarding performance risks. This result may reflect the belief of people with either acute or chronic diseases that the possibility of making medical errors with AI clinical applications would be the same as that for traditional in-person visits with physicians, even when doctors monitor the AI applications. Thus, the findings provide no solid evidence that individuals may believe that AI models and their features exhibit functional errors or technological uncertainties that endanger patient safety and lead to death or injuries. Respondents reported that any clinical encounters (ie, traditional, collaborative intelligence, or AI applications) could lead to incorrect diagnoses or wrong treatments.

Perceived Benefits

The results demonstrated significant differences in the perceived benefits according to the scenarios when using three alternative clinical encounters. Although the posthoc test did not show a significant difference among the six experimental groups, we observed that people with either acute or chronic conditions associated more benefits to direct interactions with their physicians. Among the AI options, only Acute-AI-physician (acute conditions and collaborative intelligence) showed similar responses to those given for traditional face-to-face interactions. Moreover, the scores of Acute-AI-physician on perceived risks such as privacy concerns, communication barriers, as well as regulatory and liability issues were not significantly different from those of other AI-based scenarios. Therefore, since the Acute-AI-physician scenario was associated with relatively higher benefits and nonsignificant differences in risk perceptions, we can argue that it might be a better option to start with. Accordingly, we recommend that implementing an AI-based service that physicians directly control and monitor would be an acceptable choice for patients with acute diseases.

Moreover, these results suggest that health care organizations, physicians, and AI application developers need to highlight potential AI benefits in their marketing campaigns to promote usability and the value of their AI applications, and ultimately increase the rate of usage. This argument is consistent with other studies suggesting that patients become more likely to use AI clinical applications if they believe they can improve diagnostics, prognosis, and patient management systems [95]. Specific marketing strategies in medical AI application companies and hospitals can be developed to enhance users' awareness about both human and computer intelligence. These strategies should enlighten physicians on maintaining interactions with patients while using AI clinical applications that could suggest accurate care planning, reduce health care costs, and boost health care outcomes. Thus, highlighting the performance benefits of AI, such as accuracy of diagnosis, reliability of data analysis, the efficiency of care planning, and consistency of treatments, in communication with users along with marketing materials may increase individuals' intention to at least try services provided by AI applications in health care.

Intention to Use

The results indicated that people with chronic diseases are less willing to solely use AI clinical applications or AI applications controlled by physicians. Since chronic conditions encourage patients to visit their physicians frequently to consult them about their illness signs, symptoms, and progress, they are generally more likely to prefer human-human consultations over human-computer interactions. This point highlights that patients suffering from a long-lasting disease may not be ready to use pure or partial AI clinical applications to control their chronic conditions. Therefore, health care organizations need to exercise caution when implementing these applications for chronic diseases.

Furthermore, it should be mentioned that even though one of the primary outcome variables in this study was the intention to use, we do not propose that an unconditional acceptance of

AI clinical applications is the ideal situation in health care. In contrast, we exhibit how important value-based consideration is when implementing AI applications in health care contexts. Suppose the rejection of medical AI is explained by huge and unaddressed technological, ethical, or regulatory concerns. In that case, there is not much sense in partially coping with these concerns by setting up the mandatory use of medical AI covering the entire patient spectrum. We propose that a successful rollout of AI clinical applications be managed with the knowledge and consideration of potential users' benefits and risk perceptions. There is growing interest in research about AI-centric technologies; however, individuals have not yet integrated AI applications into many aspects of their lives [96]. We can argue that the public's general technical knowledge about AI performance and how it works is still at an early stage. If AI clinical applications gained more ground in everyday care work, people would have a better perspective on the benefits and risks associated with them and actually start using them.

The Role of Training

AI clinical applications should be designed in a way to respect patients' autonomy and decision-making freedom. AI agents should not follow a coercive approach to force patients to make health-related decisions under pressure. Regulations should illuminate patients' roles in relation to AI applications so that they are aware of their position to refuse AI-based treatments where possible [97]. An important aspect that needs to be built into AI systems in health care is the transparency of AI algorithms so that the AI system does not remain a black box to the users. Technical education, health knowledge, and explicit informed consent should be emphasized in the AI implementation model to prepare patients for AI use. Training should target the patient community to ensure that the patients obtain sufficient information to make informed health decisions. Thus, if users understand the basics of AI applications, and the potential benefits and limitations they can bring to health care, they will become more willing to accept AI use to obtain improved health care delivery. Under this circumstance, users will be active partners of AI applications rather than passive AI recommendation receivers.

Limitations and Future Work

Although this study provides theoretical and practical implications, it has some limitations. First, we collected data from a sample of respondents from the United States. Care work culture and technology use are diverse among different countries. Therefore, we recommend that future studies consider subjects from other geographical locations such as other developed countries and developing countries that may not yet be implementing and using technologically advanced infrastructures in health care services (such as smart devices or AI clinical applications). Second, our study used an online survey to recruit participants digitally, and several measures were taken to provide clear definitions and scenarios. Since a self-rated sample of participants on MTurk was used, there is still a small chance that some respondents were not completely aware of AI technology and may have formed their own perceptions of the information technology artifact. Therefore, we suggest that further studies use a different method to ensure that subjects are knowledgeable about medical AI. For instance, future research can recruit informed patients who are directly referred by the providers using patient self-management tools such as wearable devices with embedded AI.

Third, due to the online data collection procedure through MTurk, we only considered respondents who could access the internet and were healthy enough to participate in an online survey. Although the MTurk pool has been recognized as an acceptable data collection means for academic research, caution should be exercised when generalizing this study's results. Future researchers may extend this study by using other data collection methods to reach out to patients. Since the experiments did not occur within a health care setting (such as a hospital), the generalizability of our findings could have been limited. Thus, it would be interesting for future studies to repeat the same experiments through simulations with treating physicians and patients suffering from acute and chronic diseases visiting a health care center (eg, a hospital).

Fourth, the lack of educational background diversity (eg, 71% of participants had higher education) and age variation (eg, 63% were younger than 40 years) of the sample may be considered a limitation for the generalizability of our results. Thus, it is recommended that future studies consider drawing samples with more representative subjects in wider age groups with various levels of education. Fifth, we used the general concept of AI, and no specific type of AI clinical application was examined. Users' perceptions may have different underlying objectives depending on the type of AI application considered. However, this study can also serve as a starting point for further empirical studies in the context of individual adoption of AI clinical applications. For instance, it would be interesting to investigate how alternative AI application brands influence risk beliefs, perceived benefits, and intention to use. Finally, we defined AI applications as tools that consumers can voluntarily choose to use for health care management. Another promising research avenue would be to examine public perspectives in other health care contexts such as when AI applications are implemented and used in hospitals and health care professionals recommend that patients start to use these applications. We also recommend a follow-up study examining users' value perceptions in the context of mandatory AI applications used for diagnosing and completing patient treatments.

Conclusions

Disruptive advances in technology inevitably change societies, communications, and working life. Technology and health care have become inseparable in recent times. One of the fundamental technological changes that could impose significant health care effects is the widespread implementation and use of AI clinical applications. AI technology is an integral element of many organizations' business models, and it is a critical strategic component in the plans for many sectors of business such as health care institutions. Implementing advanced information systems (such as AI) in health care requires an in-depth understanding of the factors associated with technology acceptance among groups of stakeholders. One of the most important stakeholders of AI clinical applications is patients. Due to the distinct characteristics of the health care sector, the

implementation of AI applications should be conducted with several necessary considerations. From the public perspective, using AI applications is a form of endorsing them. Our results highlight that there are still noticeable concerns about implementing AI clinical applications in diagnostics and treatment recommendations for patients with both acute and chronic illnesses, even if these tools are used as a recommendation system under physician experience and wisdom. Our study shows that individuals may still not be ready to accept and use AI clinical applications owing to some risk beliefs. Before implementing AI, more studies are needed to identify the challenges that raise concerns for the implementation and use of AI tools. We recommend addressing the concerns contributing to risk beliefs about using AI clinical applications as a priority for health care organizations. If privacy concerns, trust issues, communication barriers, concerns related to the transparency of regulatory standards, and liability risks are not analyzed, rationalized, and resolved accordingly, people may not use these applications. They may further view AI applications as a threat to their health care. AI application developers and health care providers need to highlight the potential benefits from AI technology and address different dimensions of concerns to justify using an AI clinical application to the public. Health care regulatory agencies need to clearly define the rights and the responsibilities of health care professionals, developers, programmers, and end users to demonstrate acceptable approaches in using AI applications in health care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Experiments and scenarios.
[DOCX File , 33 KB - jmir_v23i11e25856_app1.docx]

Multimedia Appendix 2
Demographic information of participants in the six scenarios.
[DOCX File , 28 KB - jmir_v23i11e25856_app2.docx]

Multimedia Appendix 3
Detailed Scheffe posthoc test results.
[DOCX File , 70 KB - jmir_v23i11e25856_app3.docx]

References

1. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 2019 Jan;62(1):15-25. [doi: 10.1016/j.bushor.2018.08.004]
2. Jarrahi MH. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus Horiz 2018 Jul;61(4):577-586. [doi: 10.1016/j.bushor.2018.03.007]
3. Bitner MJ, Brown SW, Meuter ML. Technology infusion in service encounters. J Acad Mark Sci 2000 Jan 01;28(1):138-149. [doi: 10.1177/0092070300281013]
4. Marinova D, de Ruyter K, Huang M, Meuter ML, Challagalla G. Getting smart. J Service Res 2016 Dec 02;20(1):29-42. [doi: 10.1177/1094670516679273]
5. Larivière B, Bowen D, Andreassen TW, Kunz W, Sirianni NJ, Voss C, et al. "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. J Bus Res 2017 Oct;79:238-246. [doi: 10.1016/j.jbusres.2017.03.008]
6. Robinson S, Orsingher C, Alkire L, De Keyser A, Giebelhausen M, Papamichail KN, et al. Frontline encounters of the AI kind: An evolved service encounter framework. J Bus Res 2020 Aug;116:366-376. [doi: 10.1016/j.jbusres.2019.08.038]
7. Chi OH, Denton G, Gursoy D. Artificially intelligent device use in service delivery: a systematic review, synthesis, and research agenda. J Hosp Mark Manag 2020 Feb 11;29(7):757-786. [doi: 10.1080/19368623.2020.1721394]
8. Huang M, Rust RT. A strategic framework for artificial intelligence in marketing. J Acad Mark Sci 2020 Nov 04;49(1):30-50. [doi: 10.1007/s11747-020-00749-9]
9. De Keyser A, Köcher S, Alkire (née Nasr) L, Verbeeck C, Kandampully J. Frontline service technology infusion: conceptual archetypes and future research directions. J Serv Manag 2019 Jan 14;30(1):156-183. [doi: 10.1108/josm-03-2018-0082]
10. Gursoy D, Chi OH, Lu L, Nunkoo R. Consumers acceptance of artificially intelligent (AI) device use in service delivery. Int J Inf Manag 2019 Dec;49:157-169. [doi: 10.1016/j.ijinfomgt.2019.03.008]
11. López-Robles J, Otegi-Olaso J, Porto Gómez I, Cobo M. 30 years of intelligence models in management and business: a bibliometric review. Int J Inf Manag 2019 Oct;48:22-38. [doi: 10.1016/j.ijinfomgt.2019.01.013]
12. Coombs C, Hislop D, Taneva SK, Barnard S. The strategic impacts of intelligent automation for knowledge and service work: an interdisciplinary review. J Strat Inf Syst 2020 Dec;29(4):101600. [doi: 10.1016/j.jsis.2020.101600]

13. Khanna S. Artificial intelligence in health – the three big challenges. Australas Med J 2013 Jun 01;6(5):315-317. [doi: 10.4066/amj.2013.1758]
14. Dreyer K, Allen B. Artificial intelligence in health care: brave new world or golden opportunity? J Am Coll Radiol 2018 Apr;15(4):655-657. [doi: 10.1016/j.jacr.2018.01.010] [Medline: 29501501]
15. Houssami N, Turner RM, Morrow M. Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer. Breast Cancer Res Treat 2017 Sep 6;165(2):273-283 [FREE Full text] [doi: 10.1007/s10549-017-4324-3] [Medline: 28589366]
16. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. J Alzheimers Dis 2015 Oct 15;49(2):407-422. [doi: 10.3233/jad-150520]
17. 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]
18. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan 7;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
19. Wilson J, Daugherty PR. Collaborative intelligence: humans and AI are joining forces. Harvard Business Review. 2018 Jun 01. URL: https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces [accessed 2021-11-20]
20. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
21. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform 2020 May;137:104072. [doi: 10.1016/j.ijmedinf.2019.104072] [Medline: 32200295]
22. Pinto Dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr 6;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
23. Gong B, Nugent JP, Guest W, Parker W, Chang PJ, Khosa F, et al. Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: a national survey study. Acad Radiol 2019 Apr;26(4):566-577. [doi: 10.1016/j.acra.2018.10.007] [Medline: 30424998]
24. European Society of Radiology (ESR). Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology. Insight Imag 2019 Oct 31;10(1):105-111 [FREE Full text] [doi: 10.1186/s13244-019-0798-3] [Medline: 31673823]
25. Broadbent E, Tamagawa R, Patience A, Knock B, Kerse N, Day K, et al. Attitudes towards health-care robots in a retirement village. Australas J Ageing 2012 Jun;31(2):115-120. [doi: 10.1111/j.1741-6612.2011.00551.x] [Medline: 22676171]
26. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, et al. Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2019 Nov 18;2(1):111. [doi: 10.1038/s41746-019-0189-7]
27. Turja T, Aaltonen I, Taipale S, Oksanen A. Robot acceptance model for care (RAM-care): A principled approach to the intention to use care robots. Inf Manag 2020 Jul;57(5):103220. [doi: 10.1016/j.im.2019.103220]
28. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020 Jul 22;20:170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
29. Lee H, Piao M, Lee J, Byun A, Kim J. The purpose of bedside robots: exploring the needs of inpatients and healthcare professionals. Comput Inform Nurs 2020 Jan;38(1):8-17. [doi: 10.1097/CIN.0000000000000558] [Medline: 31524689]
30. Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SBM, Savoldelli M, et al. Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health 2020 Jun 24;16:52 [FREE Full text] [doi: 10.1186/s12992-020-00584-1] [Medline: 32580741]
31. Zhang B, Dafoe A. Artificial intelligence: American attitudes and trends. SSRN J 2019 Jan 09:1-111. [doi: 10.2139/ssrn.3312874]
32. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
33. Hermes S, Riasanow T, Clemons EK, Böhm M, Krcmar H. The digital transformation of the healthcare industry: exploring the rise of emerging platform ecosystems and their influence on the role of patients. Bus Res 2020 Sep 11;13(3):1033-1069. [doi: 10.1007/s40685-020-00125-x]
34. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. BMJ 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]
35. Grichnik KP, Ferrante FM. The difference between acute and chronic pain. Mt Sinai J Med 1991 May;58(3):217-220. [Medline: 1875958]
36. About chronic diseases. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2020-12-05]

37. Xu J. Overtrust of robots in high-risk scenarios. 2018 Presented at: 2018 AAAI/ACM Conference on AI, Ethics, and Society; February 2-3, 2018; New Orleans, LA p. 390-391. [doi: 10.1145/3278721.3278786]
38. Wu Y, Cristancho-Lacroix V, Fassert C, Faucounau V, de Rotrou J, Rigaud A. The attitudes and perceptions of older adults with mild cognitive impairment toward an assistive robot. J Appl Gerontol 2016 Jan 09;35(1):3-17. [doi: 10.1177/0733464813515092] [Medline: 24652924]
39. Bauchat JR, Seropian M, Jeffries PR. Communication and empathy in the patient-centered care model—why simulation-based training is not optional. Clin Simul Nurs 2016 Aug;12(8):356-359. [doi: 10.1016/j.ecns.2016.04.003]
40. Spiro H. Commentary: the practice of empathy. Acad Med 2009;84(9):1177-1179. [doi: 10.1097/acm.0b013e3181b18934]
41. Charon R. The patient-physician relationship. Narrative medicine: a model for empathy, reflection, profession, and trust. JAMA 2001 Oct 17;286(15):1897-1902. [doi: 10.1001/jama.286.15.1897] [Medline: 11597295]
42. Kerasidou A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. Bull World Health Organ 2020 Jan 27;98(4):245-250. [doi: 10.2471/blt.19.237198]
43. Heinze K, Suwanabol PA, Vitous CA, Abrahamse P, Gibson K, Lansing B, et al. A survey of patient perspectives on approach to health care: focus on physician competency and compassion. J Patient Exp 2020 Dec 28;7(6):1044-1053 [FREE Full text] [doi: 10.1177/2374373520968447] [Medline: 33457544]
44. Singh P, King-Shier K, Sinclair S. South Asian patients' perceptions and experiences of compassion in healthcare. Ethn Health 2020 May 11;25(4):606-624. [doi: 10.1080/13557858.2020.1722068] [Medline: 32046500]
45. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014 Apr 9;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
46. Lu L, Cai R, Gursoy D. Developing and validating a service robot integration willingness scale. Int J Hospit Manag 2019 Jul;80:36-51. [doi: 10.1016/j.ijhm.2019.01.005]
47. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med 2018 Nov 6;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
48. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 01;27(3):491-497 [FREE Full text] [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
49. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
50. LaRosa E, Danks D. Impacts on trust of healthcare AI. 2018 Presented at: 2018 AAAI/ACM Conference on AI, Ethics, and Society; February 2-3, 2018; New Orleans, LA p. 210-215. [doi: 10.1145/3278721.3278771]
51. Bryson J, Winfield A. Standardizing ethical design for artificial intelligence and autonomous systems. Computer 2017 May;50(5):116-119. [doi: 10.1109/mc.2017.154]
52. Gupta RK, Kumari R. Artificial intelligence in public health: opportunities and challenges. JK Sci 2017;19(4):191-192.
53. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can Assoc Radiol J 2018 May 01;69(2):120-135 [FREE Full text] [doi: 10.1016/j.carj.2018.02.002] [Medline: 29655580]
54. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan 03;112(1):22-28. [doi: 10.1177/0141076818815510] [Medline: 30507284]
55. Luxton D. Should Watson be consulted for a second opinion? AMA J Ethics 2019;21(2):E131-E137.
56. Vance A, Elie-Dit-Cosaque C, Straub DW. Examining trust in information technology artifacts: the effects of system quality and culture. J Manag Inf Syst 2014 Dec 08;24(4):73-100. [doi: 10.2753/mis0742-1222240403]
57. Whittlestone J, Nyrup R, Alexandrova A, Dihal K, Cave S. Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Nuffield Foundation. 2019. URL: https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf [accessed 2021-11-20]
58. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Quart 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
59. Lee J, Kim KJ, Lee S, Shin D. Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. Int J Hum-Comput Interact 2015 Jul 09;31(10):682-691. [doi: 10.1080/10447318.2015.1070547]
60. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
61. Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. Technol Forecast Soc Change 2016 Apr;105:105-120. [doi: 10.1016/j.techfore.2015.12.014]
62. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan 7;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
63. Mitchell M. Artificial intelligence hits the barrier of meaning. Information 2019 Feb 05;10(2):51. [doi: 10.3390/info10020051]
64. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica. 2016 May. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [accessed 2021-11-20]
65. Edwards SD. The HeartMath coherence model: implications and challenges for artificial intelligence and robotics. AI Soc 2018 Mar 8;34(4):899-905. [doi: 10.1007/s00146-018-0834-8]

66. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manag 2021 Apr;57:101994. [doi: 10.1016/j.ijinfomgt.2019.08.002]
67. Stuart R, Norvig P. Artificial intelligence-a modern approach 3rd edition. Berkeley, CA: Pearson; 2016.
68. Kirkpatrick K. It's not the algorithm, it's the data. Commun ACM 2017 Jan 23;60(2):21-23. [doi: 10.1145/3022181]
69. Noble SU. Algorithms of oppression: how search engines reinforce racism NYU Press, 2018. 256 pp. Science 2021 Oct 29;374(6567):542-542. [doi: 10.1126/science.abm5861] [Medline: 34709921]
70. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans A Math Phys Eng Sci 2018 Oct 15;376(2133):20180080 [FREE Full text] [doi: 10.1098/rsta.2018.0080] [Medline: 30322996]
71. Esmaeilzadeh P. The effects of public concern for information privacy on the adoption of Health Information Exchanges (HIEs) by healthcare entities. Health Commun 2019 Sep 08;34(10):1202-1211. [doi: 10.1080/10410236.2018.1471336] [Medline: 29737872]
72. Dawson D, Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, et al. Artificial intelligence: Australia's ethics framework. Analysis and Policy Observatory. 2019. URL: https://apo.org.au/node/229596 [accessed 2020-09-23]
73. Zandi D, Reis A, Vayena E, Goodman K. New ethical challenges of digital technologies, machine learning and artificial intelligence in public health: a call for papers. Bull World Health Organ 2019 Jan 01;97(1):2. [doi: 10.2471/blt.18.227686]
74. Beregi J, Zins M, Masson J, Cart P, Bartoli J, Silberman B, Conseil National Professionnel de la Radiologie et Imagerie Médicale. Radiology and artificial intelligence: An opportunity for our specialty. Diagn Interv Imaging 2018 Nov;99(11):677-678 [FREE Full text] [doi: 10.1016/j.diii.2018.11.002] [Medline: 30473436]
75. Tizhoosh H, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. J Pathol Inform 2018;9(1):38 [FREE Full text] [doi: 10.4103/jpi.jpi_53_18] [Medline: 30607305]
76. Oakden-Rayner L. Reply to 'Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists' by Haenssle et al. Ann Oncol 2019 May 01;30(5):854 [FREE Full text] [doi: 10.1093/annonc/mdy519] [Medline: 30535295]
77. Sohn K, Kwon O. Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. Telemat Inform 2020 Apr;47:101324. [doi: 10.1016/j.tele.2019.101324]
78. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec 21;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
79. Xu N, Wang K. Adopting robot lawyer? The extending artificial intelligence robot lawyer technology acceptance model for legal industry by an exploratory study. J Manag Organiz 2019 Feb 13:1-19. [doi: 10.1017/jmo.2018.81]
80. Sánchez-Prieto JC, Cruz-Benito J, Therón R, García-Peñalvo F. Assessed by machines: development of a TAM-based tool to measure AI-based assessment acceptance among students. Int J Interact Multimed Artif Intell 2020;6(4):80. [doi: 10.9781/ijimai.2020.11.009]
81. Zhao X, Xia Q, Huang W. Impact of technostress on productivity from the theoretical perspective of appraisal and coping processes. Inf Manag 2020 Dec;57(8):103265. [doi: 10.1016/j.im.2020.103265]
82. Marakanon L, Panjakajornsak V. Perceived quality, perceived risk and customer trust affecting customer loyalty of environmentally friendly electronics products. Kasetsart J Soc Sci 2017 Jan;38(1):24-30. [doi: 10.1016/j.kjss.2016.08.012]
83. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Inf Manag 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003]
84. Lo WLA, Lei D, Li L, Huang DF, Tong K. The perceived benefits of an artificial intelligence-embedded mobile app implementing evidence-based guidelines for the self-management of chronic neck and back pain: observational study. JMIR Mhealth Uhealth 2018 Nov 26;6(11):e198 [FREE Full text] [doi: 10.2196/mhealth.8127] [Medline: 30478019]
85. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
86. Paolacci G, Chandler J. Inside the Turk. Curr Dir Psychol Sci 2014 Jun 03;23(3):184-188. [doi: 10.1177/0963721414531598]
87. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/mtp1069-6679190202]
88. Froomkin AM, Kerr IR, Pineau J. When AIs outperform doctors: the dangers of a tort-induced over-reliance on machine learning and what (not) to do about it. Ariz Law Rev 2019;61:33-99. [doi: 10.2139/ssrn.3114347]
89. Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Hum Genet 2019 Feb 22;138(2):109-124 [FREE Full text] [doi: 10.1007/s00439-019-01970-5] [Medline: 30671672]
90. 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]
91. Sharkey A, Sharkey N. Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 2010 Jul 3;14(1):27-40. [doi: 10.1007/s10676-010-9234-6]
92. Duan Y, Edwards JS, Dwivedi YK. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. Int J Inf Manag 2019 Oct;48:63-71. [doi: 10.1016/j.ijinfomgt.2019.01.021]

93. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insight Imag 2018 Oct 15;9(5):745-753 [FREE Full text] [doi: 10.1007/s13244-018-0645-y] [Medline: 30112675]
94. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 2020 Apr;104:101822 [FREE Full text] [doi: 10.1016/j.artmed.2020.101822] [Medline: 32499001]
95. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019 Jun 14;2(1):53-58. [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
96. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
97. Schiff D, Borenstein J. How should clinicians communicate with patients about the roles of artificially intelligent team members? AMA J Ethics 2019 Feb 01;21(2):E138-E145 [FREE Full text] [doi: 10.1001/amajethics.2019.138] [Medline: 30794123]

Abbreviations

AI: artificial intelligence
ANN: artificial neural network
ANOVA: analysis of variance
FST: frontline service technology
IEEE: Institute of Electrical and Electronics Engineers
IS: information systems
MTurk: Mechanical Turk (Amazon)
TAM: technology acceptance model

Edited by JMIRPE Office; submitted 18.11.20; peer-reviewed by B Fakieh, L Bouchacourt, C Gross, M Nißen; comments to author 15.03.21; revised version received 04.05.21; accepted 26.10.21; published 25.11.21.

Please cite as:
Esmaeilzadeh P, Mirzaei T, Dharanikota S
Patients' Perceptions Toward Human–Artificial Intelligence Interaction in Health Care: Experimental Study
J Med Internet Res 2021;23(11):e25856
URL: https://www.jmir.org/2021/11/e25856
doi:10.2196/25856
PMID:34842535

Review

Application Scenarios for Artificial Intelligence in Nursing Care: Rapid Review

Kathrin Seibert[1], BA, RN, MSc; Dominik Domhoff[1], BA, MA, RN; Dominik Bruch[2], BA, RN, MSc; Matthias Schulte-Althoff[3], BA, MA; Daniel Fürstenau[4,5], PhD; Felix Biessmann[6], PhD; Karin Wolf-Ostermann[1], PhD

[1]Institute of Public Health and Nursing Research, High Profile Area Health Sciences, University of Bremen, Bremen, Germany

[2]Auf- und Umbruch im Gesundheitswesen UG, Bonn, Germany

[3]School of Business and Economics, Department of Information Systems, Freie Universität Berlin, Einstein Center Digital Future, Berlin, Germany

[4]Department of Digitalization, Copenhagen Business School, Frederiksberg, Denmark

[5]Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Berlin, Germany

[6]Faculty VI - Informatics and Media, Beuth University of Applied Sciences, Einstein Center Digital Future, Berlin, Germany

Corresponding Author:
Kathrin Seibert, BA, RN, MSc
Institute of Public Health and Nursing Research
High Profile Area Health Sciences
University of Bremen
Grazer Str. 4
Bremen, 28359
Germany
Phone: 49 42121868903
Email: kseibert@uni-bremen.de

Abstract

Background: Artificial intelligence (AI) holds the promise of supporting nurses' clinical decision-making in complex care situations or conducting tasks that are remote from direct patient interaction, such as documentation processes. There has been an increase in the research and development of AI applications for nursing care, but there is a persistent lack of an extensive overview covering the evidence base for promising application scenarios.

Objective: This study synthesizes literature on application scenarios for AI in nursing care settings as well as highlights adjacent aspects in the ethical, legal, and social discourse surrounding the application of AI in nursing care.

Methods: Following a rapid review design, PubMed, CINAHL, Association for Computing Machinery Digital Library, Institute of Electrical and Electronics Engineers Xplore, Digital Bibliography & Library Project, and Association for Information Systems Library, as well as the libraries of leading AI conferences, were searched in June 2020. Publications of original quantitative and qualitative research, systematic reviews, discussion papers, and essays on the ethical, legal, and social implications published in English were included. Eligible studies were analyzed on the basis of predetermined selection criteria.

Results: The titles and abstracts of 7016 publications and 704 full texts were screened, and 292 publications were included. Hospitals were the most prominent study setting, followed by independent living at home; fewer application scenarios were identified for nursing homes or home care. Most studies used machine learning algorithms, whereas expert or hybrid systems were entailed in less than every 10th publication. The application context of focusing on image and signal processing with tracking, monitoring, or the classification of activity and health followed by care coordination and communication, as well as fall detection, was the main purpose of AI applications. Few studies have reported the effects of AI applications on clinical or organizational outcomes, lacking particularly in data gathered outside laboratory conditions. In addition to technological requirements, the reporting and inclusion of certain requirements capture more overarching topics, such as data privacy, safety, and technology acceptance. Ethical, legal, and social implications reflect the discourse on technology use in health care but have mostly not been discussed in meaningful and potentially encompassing detail.

Conclusions: The results highlight the potential for the application of AI systems in different nursing care settings. Considering the lack of findings on the effectiveness and application of AI systems in real-world scenarios, future research should reflect on a more nursing care–specific perspective toward objectives, outcomes, and benefits. We identify that, crucially, an advancement

in technological-societal discourse that surrounds the ethical and legal implications of AI applications in nursing care is a necessary next step. Further, we outline the need for greater participation among all of the stakeholders involved.

(J Med Internet Res 2021;23(11):e26522) doi:10.2196/26522

KEYWORDS

nursing care; artificial intelligence; machine learning; expert system; hybrid system

Introduction

Background

Despite a surge in funded research in the application of digital technologies toward a higher assurance of quality nursing care, in times of aging societies and skill shortages [1], the application of artificial intelligence (AI) in nursing practice is still scarce. In this context, AI can be defined as algorithms that enable learning from data sets to achieve intelligent, goal-oriented action.

Recent systematic and scoping reviews on the application of AI in nursing research (as well as in practice and emerging trends), covering original research published until October 2019, identified papers listed in medical and multidisciplinary databases. These included studies focused on machine learning (ML) methods, such as deep learning [2], or on health technologies that incorporate AI approaches themselves, such as robots or clinical decision support systems [3]. Various application scenarios have been identified, including clinical or organizational outcomes (eg, falls), admission decisions in emergency medicine, high-definition image recognition, as well as socially assistive robots or health care assistant chatbots [2,3]. In addition, recent years have seen an increase in research highlighting possibilities for the future development of AI in nursing care while underscoring the importance of collaborative, interdisciplinary research, and representative, robust data sets [2].

However, as of today, a universally accepted classification of AI subfields relevant to health, which could act as a vantage point for AI in nursing practice, is missing [4]. Prominent AI approaches include ML, expert, and hybrid systems. ML, as a method of data analysis guided by algorithms, identifies patterns in data and learns from them using different approaches [4]. This is utilized in medical diagnostics, for example [5]. Expert systems build on a knowledge base and a rule-based reasoning engine [4], which, in combination, mimic the reasoning of a human expert who would solve a complex problem by applying predefined if-then rules drawing on a specific knowledge base [6]. These systems can be found in tools that support clinical decision-making and case-based reasoning [7,8]. Hybrid systems combine different AI capabilities by integrating ML with expert systems [9-11]. AI applications aimed at determining the meaning of texts, such as clinical notes, can be found in the AI subfield of natural language processing (NLP) [4,12]. AI applications for automated planning and scheduling can be used to improve the efficiency of human procedures [4], such as generating nursing staff rosters or care-related scheduling decisions [13,14]. Applications that target image and signal processing use algorithms that typically include signal feature analysis and data classification to analyze images or data produced by movement or sound [4]. These can, for example, aim at activity and health monitoring, wound detection, or pressure injury and fall prediction or prevention [15-19].

Opportunities and Challenges for AI in Nursing Care

Turning our attention specifically to nursing care settings, the primary opportunities for applying AI include application scenarios such as decision support in complex care situations [3,18,20,21]. AI also holds great promise for supporting nurses in tasks considered to take place remotely from direct patient interactions [3,21]. High expenditures of nurses' working hours are frequently reported as being used for the documentation of care processes, with some care facilities reporting up to almost a third of daily working hours being expended for documentation processes [3]. This represents one of the many starting points from which to develop AI solutions to consistently improve nursing care processes and support nurses efficiently in their daily tasks. AI applications for the direct support of care-dependent persons and their informal caregivers are another starting point, as studies with AI approaches in different community and home care settings have shown [4-6]. This is of particular need, given that most long-term care recipients in Germany are being cared for in their own homes [22]. Until now, little knowledge on the practical relevance and applicability of AI systems with setting-specific requirements in nursing care, for example, when introduced in care processes involving persons with limited cognitive abilities, exists thus far.

Furthermore, the transformative effect of AI, resulting from its ability to change the intrinsic nature of health care delivery, is accompanied by ethical risks, namely, concerning the validity of evidence, the fairness of outcomes, and the traceability of harm caused by algorithmic activity [23]. Furthermore, although consensus on the potential of health technologies powered by AI to enhance nursing practice has been reported [3], the critical ideological and ethical nature of nursing practice still needs to be considered, and the role of decision-making, enhanced and burdened by an amplified understanding of opportunities granted by AI applications remains uncertain in the context of providing ethical and transparent nursing care [21]. To our knowledge, an extensive overview of the evidence base and status quo of research on AI for application in nursing practice, including evidence from medical and computer science databases, is missing. By identifying promising application scenarios for AI in nursing practice, such an overview contributes to the systematic enhancement of research and development for AI in nursing practice.

Objectives

This rapid review aims to synthesize the evidence base of application scenarios for AI in nursing care settings, namely, ambulatory and stationary (long-term) care, acute hospital care,

and nursing education. We also address prominent adjacent aspects within the ethical, legal, and social discourse concerning AI in nursing care by addressing the following review questions:

1. Which application scenarios for AI systems in nursing practice are reported, considering that different care settings are described in the literature?
2. What kinds of AI approaches have been researched, or are being discussed in the literature, and for which kinds of care settings?
3. What requirements or barriers have been reported for the application of AI in nursing practice?
4. Which ethical, legal, and social aspects—concerning AI and nursing—are discussed in the national and international literature?

Although our approach is broader than those of similar reviews that have focused exclusively on ML algorithms [24], it is also more broadly scoped in that it considers the ethical and regulatory context of the AI system deployed but does not focus exclusively on these aspects as other reviews have done [25].

Methods

Criteria for Considering Publications for This Review

We conducted a rapid review to identify and synthesize publications promptly [26]. A protocol describing the rationale and methods of this review was published in May 2020 [27]. This paper follows the guidelines outlined in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement [28].

We included all designs of quantitative and qualitative original research, systematic reviews, and discussion papers or essays on ethical, legal, and social aspects that address the application of AI, specifically in:

1. The support of decision or work processes in direct nursing care or,
2. The organization of nursing care processes,
3. The support of knowledge and competencies in nurses' (further) education or,
4. The support of persons in need of care (explicitly referred to as needing care) or,
5. Persons of all ages in need of support in their activities of daily living.

We included publications in the English language from 2005 onward, as we expect publications on AI to become quickly outdated and updated. Publications focusing on improving the functionality of medical diagnostic or therapeutic technologies, without clearly describing nurses as a relevant target group being affected by the application of the AI system or being directly involved in the application process, were excluded.

Types of Participants, Settings, and AI Systems

Publications had to either designate at least 1 of the following groups of persons as main users or as benefactors of an AI application:

1. Nurses or nursing students
2. Care-dependent persons or their informal caregivers (either explicitly referred to as needing care or being referred to as needing, or benefiting from physical, cognitive, or mental support).

Publications using inconclusive terms, such as *the elderly* or *health care professionals*, without further information on target groups of users or benefactors, were also assessed for inclusion.

Care settings encompassed ambulatory and stationary long-term as well as acute outpatient and hospital care (including rehabilitation facilities) and nursing education settings. Studies assessing care in community settings or assessing the populations mentioned above, but in a laboratory setting, were also included.

As there is no conclusive definition of specific AI abilities or subfields that are relevant for health [4] or nursing care as of yet, all types of AI systems or approaches, ranging from clearly stated types (ML, expert system, hybrid system), to any type of approach combining ML and an expert system and algorithms, across to rather vague descriptions, such as *smart system* or *AI in health care*, were deemed as eligible.

Search Methods for Identification of Studies

We searched the following databases in June 2020: PubMed, CINAHL (including Embase), Association for Computing Machinery Digital Library, Institute of Electrical and Electronics Engineers Xplore, Digital Bibliography & Library Project, computer science bibliography, and Association for Information Systems Library. In addition, we searched digital libraries of leading conferences identified through expert consensus within the study team. These conferences were specifically the Association for the Advancement of AI conference, the Association for Computational Linguistics Conference, the Conference on Computer Vision and Pattern Recognition, the International Conference on Machine Learning, the International Joint Conferences on AI Organization, the conference of the Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining, the Conference on Neural Information Processing Systems, the International Conference on Principles of Knowledge Representation and Reasoning, the Conference on Uncertainty in AI, the International Conference on Autonomous Agents and Multiagent Systems, and the European Conference on AI. The search strategy based on the block-building approach [29], combined terms for *nursing* and *artificial intelligence* and their respective synonyms. If applicable, we searched the titles, abstracts, and all fields of publication. In the first step, single terms for each block were searched. Second, all terms of a single block were combined using the Boolean operator *OR*. We initially deemed publications in German or English to be eligible. As no publications in the German language fulfilled the inclusion criteria, we focused only on the English. Finally, the results from the second step were combined for the 2 blocks using the Boolean operator AND. The hits were recorded for each step. To circumvent imprecisions between concepts described in the titles or abstracts, regarding the components of the search strategy, and to identify a large number of potentially eligible publications, we developed a preferably sensitive search strategy. Multimedia Appendix 1 contains the search strategy,

search terms, and the number of hits for all databases and conference libraries.

Data Collection and Analysis

Selection of Publications, Data Management, and Extraction

Two review authors independently screened all the titles and abstracts. Full texts were screened by a single person. Discrepancies were resolved through discussion or by referral to a third review author. Citations identified by the third search step described above were exported to an EndNote library after excluding duplicates. Titles and abstracts were screened using the web-based resource, Rayyan [30]. Full text screening was conducted in EndNote and documented using a spreadsheet program. Data extraction was conducted by a single reviewer and included the following data for all publications:

1. Author, year, country of origin
2. Setting
3. Target group of users or benefactors
4. Methods used or addressed
5. Purpose of the AI application

Furthermore, we extracted information on study design, type of data sets used, number of participants, outcomes assessed, results, and reported requirements or barriers for the application of AI for a subsample of studies that we considered as studies that incorporated real-world settings. For studies focusing on research of a more basic nature and describing laboratory scenarios or which used pre-existing data sets, either without transfer of results to real-world nursing scenarios or without evaluation of real-world outcomes, or focused on algorithm qualities or proof-of-concept studies, no information on results were extracted.

Assessment of Risk of Bias and Level of Evidence

As the rating of the effectiveness of AI applications in nursing care was not a primary research interest of this review, we did not assess the risk of bias of the results within the original research studies. To map the advancement of research regarding reliability, external validity, and generalization of results, a level of evidence (LOE) was assigned to each publication. We used established evidence-based nursing and evidence-based medicine hierarchies [31,32] and ranked LOEs from level I (highest evidence) to level VII (lowest evidence), as shown in Textbox 1. As we did not assess the risk of bias, the characteristic *well designed* in the LOE description is enclosed in brackets. Studies using a nonrandomized control group design, in which one group of participants did not receive an AI-supported intervention, or where a before-and-after design was implemented, were assigned to level III. Publications providing an overview without using a systematic review design were assigned to level VII. Publications not reporting results obtained by a specific research design were then labeled as *concept only* and were not assigned an LOE.

Textbox 1. Level of evidence rating categories.

Level of evidence

- Level I
  - Evidence from a systematic review or meta-analysis of all relevant randomized controlled trials (RCTs) or evidence-based clinical practice guidelines, based on systematic reviews of RCTs, or of 3 or more RCTs of good quality that have similar results
- Level II
  - Evidence obtained from at least 1 (well-designed) RCT
- Level III
  - Evidence obtained from (well-designed) controlled trials without randomization (eg, quasi-experimental)
- Level IV
  - Evidence from (well-designed) case-control or cohort studies
- Level V
  - Evidence from systematic reviews of descriptive and qualitative studies (metasynthesis)
- Level VI
  - Evidence from a single descriptive or qualitative study
- Level VII
  - Evidence from the opinion of authorities or reports of expert committees
- No applicable level
  - Concept only

Analysis and Synthesis

Publications were grouped into basic research studies (category *basic or experimental*) or those incorporating real-world scenarios. The country of origin was coded into a country code, as defined in ISO 3166-1; it refers to the country in which the analyzed data were generated. We classified studies as either directly addressing nurses, care dependents, patients, or informal caregivers as being the main users or benefactors of the AI system. An AI system can solve complex problems that have been previously reserved for humans. This is done by breaking these problems into a number of simple prediction tasks [33]. We coded the types of AI systems and application contexts for each publication on the basis of the categories given in Wahl et al [4], which we expanded after determining the final sample of publications to be included. The category *Type of AI Approach* comprises the codes *machine learning* and *expert system*, as defined above. In addition, we also considered *hybrid systems*, defined as a combination of expert systems with ML [9-11]. Studies using deep learning approaches have been included in ML. AI systems can be defined as self-training structures of ML predictors, which automate and accelerate human tasks, and consist of *domain structure*, *data generation*, and a *general purpose prediction algorithm* [33]; information on the domain structure, which needs to be attributable to the nursing care context, was mandatory for inclusion. Studies lacking information on the data generation, as well as the prediction algorithm dimension, using rather generalized terms, were categorized as not specified in the category *Type of AI Approach*.

The application context category is also derived from Wahl et al [4] and comprises *automated planning and scheduling*, *image and signal processing*, and *NLP*. Both categories also entail codes for unclear and nonspecific information or restricted applicability. Originating from the data extracted for the purpose of the AI application, we inductively developed codes for the *setting* category and 22 codes for the *purpose* category that summarize the domain of health or nursing activity affected by the AI system (eg, nurse rostering and scheduling, tracking or monitoring of activity and health tracking, falls or quality of life, and well-being of caregivers). In addition, we inductively derived 7 codes for a more generalized *application scenario* (support of direct nursing care, support of the care organization, support of independent living care-dependent people, health of the caretaker, formal and informal education, risk estimation and prevention, etc). Systematic reviews and other types of publications were coded as described above if possible, or rated as not applicable for some categories.

Study characteristics and target groups of users or benefactors are descriptively summarized and displayed in tables and figures. To answer the first and second research questions, we descriptively summarized the categories *purpose*, *application scenario*, *type of AI approach*, and *application context* in relation to the *setting* category. The results will be summarized, as well as differentiated, for studies considered to be of a more basic nature, such as laboratory experiments or proof-of-concept papers (category *basic or experimental*) and real-world scenario studies (category *real-world setting*) and displayed in tables or figures. To answer the third research question, we narratively summarized the requirements and barriers reported in real-world scenario studies, as well as systematic reviews or publications focusing on the ethical, legal, and social implications (ELSIs) of AI in nursing care. The latter also provides the basis for answering the fourth research question in the form of narrative synthesis.

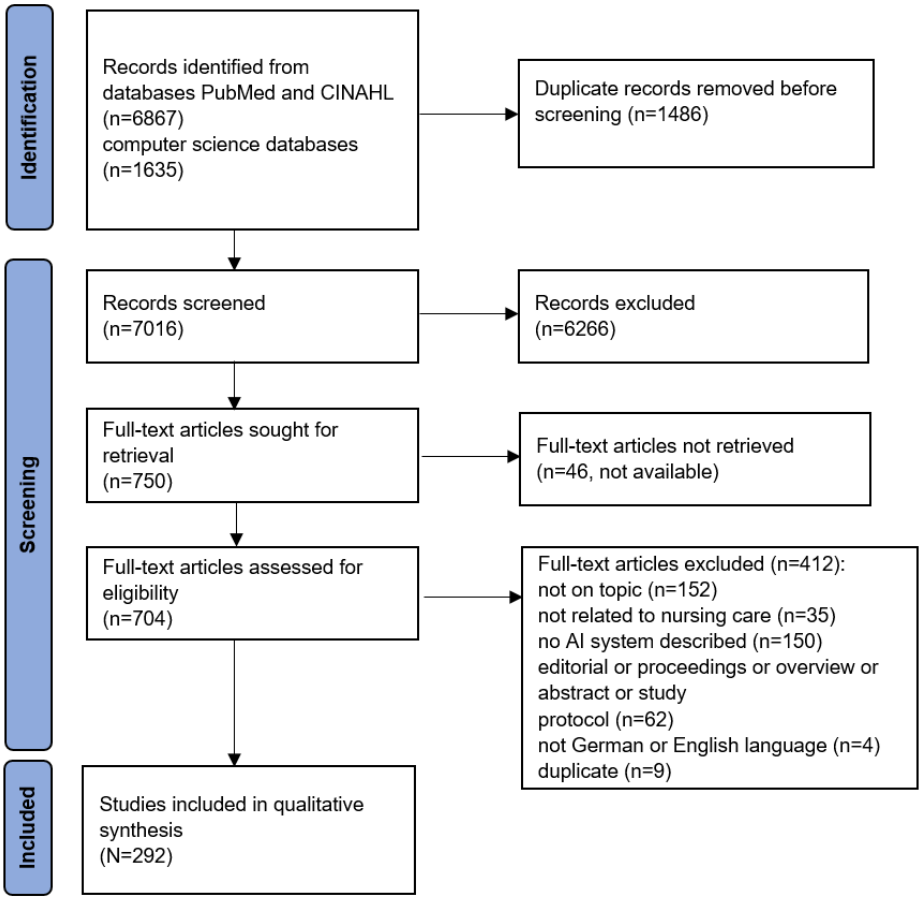
Results

Included Publications

Overview

Searches performed in databases for nursing and health sciences yielded 6867 matches. Databases containing publications from computer science publications added an additional 1635 matches. The handling of the included publications and the numbers of included and excluded records are depicted in Figure 1. In the first step, we eliminated duplicate records (n=1486), resulting in 7016 publications proceeding into the screening of their titles and abstracts, which led to the further exclusion of 6266 publications. For the remaining 704 available publications, the full texts were screened, and a further 412 publications were excluded in this step, leaving 292 publications to be incorporated in this review (Multimedia Appendix 2 [7-13,15-19,21,34-310] describes an overview of all 292 references and the selected characteristics for these included publications).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [34] flowchart of the publication screening process and study selection. AI: artificial intelligence.



Characteristics of Included Studies

Publication Date, Country, and Publication Language

Of the 292 studies, 155 (53.1%) were published between 2016 and June 2020, with the remaining 137 (46.9%) originating from 2005 to 2015. The included studies used data generated in 39 countries, which in most cases corresponded with the country affiliation of the first author. The 10 countries with the most publications were the United States (n=72), Japan (n=45), Canada (n=23), China (n=16), Taiwan (n=15), the United Kingdom (n=11), Australia (n=10), India (n=9), Spain (n=9), South Korea (n=8), and Germany (n=8). All studies were published in English.

Research Setting

We classified 83.2% (243/292) of studies as basic or experimental and 11.6% (34/292) as studies in real-world settings. In addition, 8 scoping or systematic reviews, 6 publications on ethical, legal, and social aspects, and 1 survey on AI in nursing practice were included.

Level of Evidence

The LOE assigned most often was level VI (evidence from a single descriptive or qualitative study), which applied to 77.1% (225/292) of studies. An evidence level of III (evidence obtained from well-designed controlled trials without randomization) or higher was assigned to 2.1% (6/292) of studies. Table 1 shows the number of studies assigned to each level of the evidence category.

Table 1. Numbers of publications by level of evidence and research setting (N=292).

Level of evidence	Basic or experimental (n=243), n (%)	Real-world setting (n=34), n (%)	Other (n=15), n (%)	Total (N=292), n (%)
Level I	0 (0)	0 (0)	1 (6.7)	1 (0.3)
Level II	0 (0)	1 (2.9)	0 (0)	1 (0.3)
Level III	0 (0)	4 (11.8)	0 (0)	4 (1.4)
Level IV	7 (2.9)	7 (20.6)	0 (0)	14 (4.8)
Level V	0 (0)	0 (0)	7 (46.7)	7 (2.4)
Level VI	199 (81.9)	21 (61.8)	5 (33.3)	225 (77.1)
Level VII	4 (1.6)	0 (0)	2 (13.3)	6 (2.1)
No applicable level	33 (13.6)	1 (2.9)	0 (0)	34 (11.6)

Beneficiaries and Setting

We found that 46.2% (135/292) of publications specifically addressed care-dependent persons as the target group of the proposed or examined AI solutions, 39.4% (115/292) targeting nurses, and 9.6% (28/292) stating informal caregivers as the target group. Two or more of the aforementioned groups were addressed in 49 studies, and all of them in 6 publications. In addition, 23.3% (68/292) of the publications did not state either of the 3 groups as their primary target group. These studies frequently proposed AI approaches with nurses targeted as potential beneficiaries among other health care professionals.

Hospitals are the most prominent research setting, followed by independent living at home, with nursing homes, ambulatory long-term care, and outpatient health care being less frequently addressed (Table 2). Other settings, including the community, rehabilitation, daycare, and education facilities have been the subject of only a few studies. Multiple settings were the subject of 10.6% (31/292) of the publications, and 11.3% (33/292) did not state any setting. Studies employing a real-world setting also predominantly focused on hospitals. Other settings were only referred to infrequently.

Table 2. Numbers of publications by application and research setting (N=292).

Setting	Basic or experimental (n=243), n (%)	Real-world setting (n=34), n (%)	Other (n=15), n (%)	Total (N=292), n (%)
Hospital	70 (28.8)	13 (38.2)	4 (26.7)	86 (29.8)
Independent living	64 (26.3)	0 (0)	2 (13.3)	66 (22.6)
Nursing home	21 (8.6)	12 (35.3)	0 (0)	32 (11.3)
Ambulatory long-term care	11 (4.1)	6 (17.6)	0 (0)	18 (5.8)
Outpatient health care	10 (4.1)	0 (0)	0 (0)	10 (3.4)
Community	6 (2.5)	1 (2.9)	1 (6.7)	9 (2.7)
Rehabilitation	2 (0.8)	0 (0)	0 (0)	2 (0.7)
Daycare	0 (0)	1 (2.9)	0 (0)	1 (0.3)
Education facility	0 (0)	1 (2.9)	0 (0)	1 (0.3)
Multiple	26 (10.7)	0 (0)	5 (33.3)	31 (10.6)
N/A[a]	3 (1.2)	0 (0)	0 (0)	3 (1)
Not stated	30 (12.3)	0 (0)	3 (20)	33 (11.3)

[a]N/A: not applicable.

Type and Subtype of AI Approaches

Considering the type of AI approach, we found the vast majority (228/292, 78.1%) of studies employing ML approaches. Rule-based expert systems were used in 11.6% (34/292) of the publications, whereas hybrid systems were used in only 3 studies. For the remainder of the publications, the specific AI approach used was either not identifiable or this attribute was not applicable, mostly because of publications not employing specific AI approaches in basic or applied research. Studies incorporating real-world settings made use of ML approaches comparably in 71% (24/34) of cases, whereas expert systems were only covered in 2 studies and hybrid systems not at all (Table 3).

Most AI approaches have been described as solutions for image and signal processing (178/292, 60.9%), that is, the processing of large amounts of signals, such as audio and video data, for feature analysis and data classification [4]. AI approaches have been used for automated planning and scheduling. This category

entails approaches used to organize and prioritize activities and that "can be used to improve the efficiency of human procedures" [4]. Studies have focused less often on the processing of human language (NLP). Image and signal processing were also performed in most studies conducted in a real-world setting with automated planning and scheduling, and research on NLP has been less frequently reported (Table 4).

Table 3. Numbers of publications by type of artificial intelligence (AI) system and research setting (N=292).

Type of AI system	Basic or experimental (n=243), n (%)	Real-world setting (n=34), n (%)	Other (n=15), n (%)	Total (n=292), n (%)
Machine learning	197 (81.1)	24 (70.6)	7 (46.7)	228 (78.1)
Expert system	29 (11.9)	4 (11.8)	1 (6.7)	34 (11.6)
Not specified	12 (4.9)	6 (17.6)	7 (46.7)	25 (8.6)
Hybrid system	5 (2.1)	0 (0)	0 (0)	5 (1.7)

Table 4. Numbers of publications by subfield of artificial intelligence (AI) and research setting (N=292).

Subfield of AI	Basic or experimental (n=243), n (%)	Real-world setting (n=34), n (%)	Other (n=15), n (%)	Total (N=292), n (%)
Image and signal processing	155 (63.8)	19 (55.9)	4 (26.7)	178 (61)
Automated planning and scheduling	59 (24.3)	14 (41.2)	1 (6.7)	74 (25.3)
Natural language processing	26 (10.7)	1 (2.9)	0 (0)	27 (9.2)
Not specified	3 (1.2)	0 (0)	10 (66.7)	13 (4.5)

Purpose of AI Application

The areas of support for nursing care targeted by the AI approaches are shown in Table 5. With regard to the intended effects from the described AI approaches, 47.6% (139/292) of the studies focused on the support of the direct, immediate process of care. The support of the organization of care services and the support of care-dependent people themselves, as well as risk estimation and prevention, are further prominent purposes. Risk estimation and prevention potentially pose a cross-sectional topic, where the type of support manifests at multiple levels. The health of the caregiver and education were addressed in only a few cases. For studies in real-world settings, risk estimation or prevention, support of direct care, and support of care organization were each a focus of 29% (10/34) of the studies, whereas the support of care-dependent people and education did not play a prominent role, and no research conducted in a real-world setting focused on the health of caregivers.

Table 5. Numbers of publications by area of support and research setting (N=292).

Area of support	Basic or experimental (n=243), n (%)	Real-world setting (n=34), n (%)	Other (n=15), n (%)	Total (N=292), n (%)
Support of direct care	125 (51.4)	10 (29.4)	4 (26.7)	139 (47.6)
Support of care organization	42 (17.3)	10 (29.4)	0 (0)	52 (17.8)
Risk estimation or prevention	32 (13.2)	10 (29.4)	0 (0)	42 (14.4)
Support of care-dependent people	33 (13.6)	3 (8.8)	3 (20)	39 (13.4)
N/A[a]	5 (2.1)	0 (0)	6 (40)	11 (3.8)
Health of caretaker	4 (1.6)	0 (0)	0 (0)	4 (1.4)
Various	1 (0.4)	0 (0)	2 (13.3)	3 (1)
Education	1 (0.4)	1 (2.9)	0	2 (0.7)

[a]N/A: not applicable.

A more detailed summary of the purpose of the AI approaches is presented in Table 6. The most prominent purpose was activity and health tracking (monitoring or classification) in 30.1% (88/292) studies. Care coordination and communication are frequent topics, which, among others, include AI approaches classifying information in nursing documentation, supporting decision-making, and yielding information for coordination and continuity of care. Fall detection, fall prevention, and fall risk classification are also frequently mentioned purposes for topics in AI. In contrast to falls, other mobility-related aspects were of lesser interest and were mentioned in only a few studies. Further purposes with a high degree of specificity are the recognition, classification, reduction of alarms, and risk prediction and classification of pressure ulcers. Addressing nurse rostering or scheduling problems was the purpose of an AI solution in 4.1% (12/292) studies.

Table 6. Frequencies of stated purposes (monitoring, tracking, classification, prediction, and support) of artificial intelligence solutions (N=292).

Purpose	Frequency, n (%)
Activity and health	88 (30.1)
Care coordination and communication	53 (18.2)
Falls	36 (12.3)
Nursing assessment or care needs assessment	21 (7.2)
Alarms	14 (4.8)
Nurse rostering or scheduling	12 (4.1)
Pressure ulcers	11 (3.8)
Social integration and participation	10 (3.4)
Parenteral or enteral nutrition and fluid intake	7 (2.4)
Quality of life and well-being of caregivers	6 (2.1)
Mobility, other	5 (1.7)
Speech	5 (1.7)
Distribution of medication	3 (1)
Wound management (excluding pressure ulcers)	3 (1)
Bladder control	2 (0.7)
Infection control	2 (0.7)
Respiratory care or weaning	2 (0.7)
Clinical education	1 (0.3)
COPD[a] care	1 (0.3)
Digestion management	1 (0.3)
Pain assessment or management	1 (0.3)
N/A[b]	8 (2.7)

[a]COPD: chronic obstructive pulmonary disease.

[b]N/A: not applicable.

Subsample of Studies in Real-world Settings

Overview

We classified 34 publications as studies that employed real-world settings. Multimedia Appendix 3 [13,18,19,34-58,63,69,82,95,100,310,312] summarizes the characteristics of these publications. The data used originated from the United States in 12 studies, from Canada in 5 studies, and from Spain in 3 studies. Australia, Germany, and Japan contributed 2 studies each to the subsample, and the remaining single studies used data from Brazil, Finland, Greece, Hong Kong, Ireland, Italy, and Singapore. A Saudi Arabian survey on health care employees' perceptions of the use of AI applications that involved 121 nurses as participants [34] was also included in the subsample. In the 27 studies that reported on the number of participants, sample sizes ranged from small samples including <10 people [35,36] to large data sets holding information from >200,000 individuals [13,37]. Of the 31 studies reporting more details on participant characteristics, data from patients in hospitals were analyzed in 13 studies, of which 5 focused on pediatric or adult intensive care unit (ICU) patients [13,38-40,312]. Residents of long-term care institutions were included in 12 studies that sometimes also included caregivers and other health professionals [35,41-44]. People with dementia or cognitive impairment were included in 3 studies [36,42,45]. Home care clients or community-dwelling elderly were included in 6 studies [46-51,313], 3 studies specifically focused on nurses or nursing students [34,52,53], and 1 study targeted elderly people at a daycare facility [54]. The more or less detailed reporting of heterogeneous study designs included experimental designs, field experiments [54], real-life use-cases [55], case studies with a single subject design [36], cross-sectional and longitudinal observational designs as well as comparative designs, and an economic evaluation nested within a cluster-randomized controlled trial [18,45], and different mixed methods designs (such as the studies by Amato et al [35], Ala-Kitula et al [46], and Alwan et al [47]). However, some studies did not state a specific design. In that case, they were classified according to the nature of the reported results (such as observational data).

Reported Effects Referring to Clinical or Organizational Outcomes

Of the studies in the subsample reporting results in varying degrees of detail, 22 studies reported effects in terms of algorithm eligibility or technological functionality. For example,

Chen et al [42] developed a detector for elopement behavior in dementia care units on the basis of a hidden Markov model and concluded that the system may reduce the risk of actual unwitnessed elopement, thus preventing negative consequences of elopement, but did not report on the longitudinal implementation of the detector or changes in elopement rates or nursing work processes. In contrast, no additional benefit by applying AI technology for a use-case aimed at gathering personal health data from home monitoring sensors, activity trackers, national electronic health records, and previous home care reports to evaluate the home care need and its availability in real time was reported by Ala-Kitula et al [46]. Results for outcomes that we considered to be of a more clinical or organizational nature are reported in 12 studies and highlight the real-world effect of the contribution of AI approaches to nursing care. Of those studies, 7 were conducted in long-term care facilities, 3 in the hospital setting, and 1 study in a daycare facility or an educational setting. Outcomes mainly target some form of physical activity, movement, or response but also, among others, length of stay (LOS), mortality, pressure ulcers, and handwashing skills.

Bajo et al [52] tested the ability of a multiagent architecture for geriatric residences to provide efficient working schedules by utilizing ML in a case-based reasoning approach. In a sample of 10 nurses, the time spent on supervision and control tasks as well as the time spent on attending to false alarms was reduced, whereas the time for direct patient care increased during the observation period of 6 months [52]. Another multiagent system to enhance assistance and health care for patients with dementia living in geriatric residences with reasoning and planning mechanisms was introduced by Tapia et al [56]. The application of the system led to a reduction in the average number of minutes spent by nurses on the monitoring of residents from more than 150 daily minutes (before implementation) to approximately 90 daily minutes (after implementation). In addition, the number of nurses working simultaneously before and after the implementation of the system reduced, and unauthorized access to restricted zones of the residence was detected almost twice as often after implementation [56]. Tang et al [43] developed a cloud-based nursing care planning system and applied case-based reasoning and text mining to facilitate decision-making of nurses responsible for admissions in a nursing home. In an observational study lasting 6 months, the efficiency of nursing care plan formulation and the response time in handling new applications increased, whereas the number of revisions of the care plan decreased. The time waiting for supporting documents reduced from 24 hours before the implementation of the system to 6.75 hours after the implementation, and the time spent searching for health care information reduced from 90 to 20 minutes, whereas the adoption of traditional health care services increased, and the residents' complaint rate decreased [43]. Xiong et al [44] examined the use of a scalable AI-enabled camera monitoring system to detect and record falls and notify nurses to perform video review of the incident immediately after each fall of residents with dementia in residential care facilities. Compared with a control group of residents who also experienced falls but were not monitored by the system, relative reductions of emergency medical team visits and emergency department visits of 75% (emergency medical team visits: P=.001) and 80% (emergency department visits: P=.003), respectively, were observed. [44]. Cho et al [18] developed a decision support intervention using a Bayesian network model to predict hospital-acquired pressure ulcers and assessed its effectiveness on the prevalence of ulcers and ICU LOS as well as on the user adoption rate and attitudes in a controlled trial. Patients in the intervention group had a decreased risk of developing hospital-acquired pressure ulcers (odds ratio 0.1; P=<.001) and a shorter ICU LOS (odds ratio 0.67; P=<.001), whereas nurses expressed favorable attitudes toward using the system [18]. Evans et al [57] developed an expert system to identify early signs of physiological deterioration in hospital patients and conducted a longitudinal evaluation of its impact on ICU transfer rates, medical emergency team calls, and mortality. During the 1-year intervention, ICU transfers and medical emergency team calls increased significantly and mortality decreased significantly when compared with the preintervention year for patients on a medical and oncology floor, whereas no significant increase was found for patients on a non-ICU surgical trauma floor that were younger and had fewer comorbidities than patients on the medical and oncology floor [57]. Yamamoto et al [53] used ML to evaluate handwashing skills in nursing students and reported that, when comparing students 3 months after their last training and beginners in handwashing, handwashing skills were almost identical, indicating the need to update practice handwashing beyond initial training. Viswanathan et al [36] observed residents of a long-term care facility with mild-to-moderate cognitive impairment while using an intelligent wheelchair and reported lower mean frontal collision with objects compared with using the wheelchair without the avoidance module preventing wheelchair motion toward nearby obstacles being activated while observing large differences in the users' collision avoidance ability. Zampieri et al [38] used ML to assess whether ICU staffing features are associated with improved hospital mortality, ICU LOS, and duration of mechanical ventilation and identified 3 clusters with, for example, the extent of nurse autonomy as a distinguishing feature and the cluster with the highest nurse autonomy exhibited the best outcomes, with lower adjusted hospital mortality and shorter ICU LOS for patients surviving to ICU discharge and shorter durations of mechanical ventilation. For studies including robotic systems, Mervin et al [45] found marginally greater values in terms of incremental cost per Cohen-Mansfield Agitation Inventory Short Form point averted from a provider's perspective for a plush-toy alternative than an emotional robot seal, but deemed the robot seal a cost-effective psychosocial treatment option for reducing agitation in people with dementia in long-term care as well, as costs are much lower than those estimated for psychosocial group activities and sensory interventions. Matsuyama et al [54] introduced a communication robot to activate and improve group communication and observed an increase in participation in terms of the frequency of smiles and answers in response to the robot system. Carros et al [41] explored, among other outcomes, stakeholders' attitudes, social and organizational practices, and expectations of the Pepper robot in individual and group-based performances and revealed the potential for humanoid robots working in

nursing homes, as well as the necessity of a person in control of the robot acting as a moderator.

Requirements and Barriers

Requirements or challenges and barriers for the application of AI in nursing care outside of technological infrastructure or reliability, precision, and validation of data have been reported in 6 studies, including real-world scenarios [18,35,37,41,58,314] and 4 reviews [59-62]. Requirements included compliance with the EU General Data Protection Regulation and preferences of target users concerning usability and complexity, and also requirements stemming from implementation in specific care settings [35,60]. Furthermore, providers need to concern themselves with their capacity, ability, and willingness to generate data inputs required to achieve high accuracy, in contrast to the clinical burden of false positive or negative results [58]. In addition, the quality of administrative databases should not be affected by AI implementation in the care context [58]. Inclusion of caregivers, user engagement, and commitment to further the participation of older adults in the development and testing of AI systems as well as successful implementation are required [18,41]. Reported challenges and barriers target accuracy of recognition, integration with sensor networks, privacy, security, human–machine interaction, and cognition impairment of users, acceptance, and costs [59,61]. The physical appearance and programmed behavior of hardware hosting AI systems when presented in a humanoid form may seem confusing, unpredictable, or frightening and limit the interpretability of the system's action for nursing home residents [41]. In addition, the appearance of sensors when they are nondisposable and bulky may pose a burden to caregivers [63]. Underreporting of relevant events and scarce public availability of databases holding sufficient data and information to compare one's results, as well as limitations to data sets due to regional data protection laws, constitute barriers to the accuracy of algorithms and the external validity of results [37,62].

Ethical, Legal, and Social Aspects

We identified 6 publications [7,21,64-67], specifically focusing on the ethical, legal, and social aspects of AI in nursing care. In addition, 7 publications [35,41,57,59,60,68,69], which were either reviews or studies including real-world settings, addressed selected ELSI aspects when discussing results or limitations of their work. Recurring aspects were *consent* (of care dependents or nurses), *data privacy and safety*, *acceptance* and *implications for work processes and workforce*, such as lack of human interaction and communication skills or the fear of replacement of nurses by technology and the implications arising from choosing humanoid designs for hardware hosting intelligent technologies [41,67]. Peirce et al [21] focused on relevant ethical, legal, and social aspects of nursing as a profession and highlighted implications arising from the type of data used, such as possible sampling bias, correlational false positives, and hidden discrimination, as well as the values and interests of companies building huge data sets that should be kept in mind. They pointed out the importance of nurses' understanding of the underlying motivations and goals for creating algorithms as well as the learning mechanisms and potential to mediate, as AI-generated knowledge should not be regarded as universally valid and the potential of algorithms to limit nursing actions and cause the loss of human dignity should be regarded of utmost importance within the discourse on AI in nursing [21].

Discussion

Principal Findings

The results of this rapid review explicate the application scenarios for AI systems in nursing care. Hospitals, followed by independent living at home, were the most frequently investigated settings, whereas nursing homes and ambulatory long-term care were less often examined. The vast majority of studies applied ML, whereas expert and hybrid systems were used only in about every 10th publication. This implies that the current instance of AI is mainly ML driven. For instance, Taddy [33] described the evolution of ML toward status as a general purpose technology and, consequently, as the main driver of the current rise in AI. More than half of the publications focused on image and signal processing and one-third on automated planning and scheduling, whereas NLP appeared in less than 1 in 10 publications.

In the context of direct nursing care, AI is used to organize care processes and support care-dependent people or family caregivers through tracking, monitoring, or classifying activities and health data. This was followed by applications to support care coordination or communication, as well as nurse rostering and scheduling. Detecting, classifying, and preventing falls, as well as recognizing, classifying, and reducing alarms, and predicting and classifying pressure ulcers were further purposes of introducing AI to nursing care. Only a few publications went beyond proof-of-concept studies or laboratory experiments and applied AI in real-world scenarios. Few studies have assessed the effects of AI on clinical and organizational outcomes. In addition to technical or computational requirements, further requirements concerning the specific context of nursing care are scarce and mainly tackle overarching topics, such as data privacy, safety, and acceptance. The same holds for ELSIs, which, for instance, have not been reflected or discussed in most studies using real-world scenarios.

Most studies describe AI applications in hospital settings, particularly ICUs. This may be attributed to the availability of such data. Besides electronic medical, nursing, or health record data, real-time sensor data on vital parameters are more frequently available from ICUs than from regular wards, facilitating a multidimensional approach, which is, for instance, being used to classify risks or identify care needs. This finding is in line with a previously reported increase in the publication rate of studies using ML to analyze routinely collected ICU data [70]. The availability, quality, and quantity of data might also be limited due to differences in digitalization activities in specific care settings, as well as the sufficient inclusion of study participants and duration observation periods to generate large data sets from sensor data. Furthermore, the heterogeneity of different data sets complicates the use of data for AI development.

Considering the development of digital technologies in general, nurses themselves have reported that they feel that regular

hospital wards or long-term care settings are being considered too little [315], which is consistent with our results. It should be noted that some of the included application scenarios cannot be attributed unambiguously to the nursing domain. Although an impact on nurses and patient care is evident when trying to reduce false alarm rates in monitoring [71,316] or to improve mechanical ventilation and sedative dosing processes [72], there is a blurred line between AI systems to support medical diagnostics and therapy, and AI systems to support nursing care. Considering the variety of applications for tracking, monitoring, and classifying health and activity, nurse rostering and scheduling, or detecting falls or fall risks, it is remarkable that so few studies went beyond testing the efficacy of AI approaches. This points to a gap in the existing evidence regarding the effectiveness of AI in real-world scenarios. This is also reflected in the results of the LOE ranking, with only a few studies using designs that could test effects on real-world nursing or on patient outcomes.

In addition, the explicit operationalization of nursing tasks or care processes, and desired clinical, psychosocial, or organizational outcomes, has not been addressed in most of the included studies. On the one hand, this might point to undiscovered possibilities for AI support in nursing care. On the other hand, it might limit the perceived benefit of AI in nursing care and, subsequently, the participation of providers, nurses, care dependents, and family caregivers in developing (as well as sustainably integrating) AI in care processes and everyday activities. Integration of AI in nursing care might also be limited by the lack of a sound description of outcomes, benefits, or values, which will influence the adoption or nonadoption of technologies in nursing care [315,317]. Our review points to a gap in published research on possible application scenarios for AI in nursing care on the one hand, and on the availability of evaluation results regarding already implemented AI systems on the other. This raises the presumption that such evaluations have so far been of less interest. This is particularly troublesome because some studies suggest little or no extra benefit of using AI when compared with alternative or existing solutions [45]. Concerning nurses' need for technologies providing enhanced technological support of direct nursing care tasks to reduce physical burdens and mental stressors [315], there seems to be room for research on AI-sensitive outcomes in nursing care.

Concerning the requirements and barriers for AI in nursing care, we expected to find topics such as data quality and access, as well as factors associated with measuring primary data and obtaining and sharing routine data, more frequently reflected in the included publications. However, only a few studies have addressed these concerns. Most requirements and barriers mirrored topics that are not only relevant for AI systems, but also for digital technologies in nursing and health care in general, such as data privacy, safety, and user acceptance [318]. On the one hand, this could indicate that there are few nursing-specific requirements or barriers to consider. However, this seems unlikely given the heterogeneous origins of the included studies, which have been conducted in different societal and health systems, including, for instance, different regulations on data protection or storage. In contrast, the lack of data and access-related factors could be attributed to the fact that the descriptive or conceptual nature of most studies led to authors addressing requirements or barriers less frequently.

The ELSIs discussed in the included publications addressed prominent topics in the discourse on the use of technology in health care, such as data privacy and protection, consent, acceptance, and implications for communication [20,25]. These aspects were not addressed in most studies in the subsample of studies, including real-world settings. Only one publication focused specifically on the ethical implications arising from the knowledge generated by AI systems in the context of nursing care. Other ethical principles incorporated in existing AI guidelines, such as trust, sustainability, justice, and fairness [25], are covered superficially if at all. Even though the limited uptake of ELSI aspects in published research might be biased by the fact that the remaining publications were not screened for ELSI aspects, there seems to be room for researchers to incorporate the discussion of ELSI aspects in their work, contributing to building trustworthy and trusted AI solutions [319]. Furthermore, publications describing stakeholder processes, surveys, interviews, or focus groups involving care dependents or nurses and accessing their perspective on AI, were underrepresented in our sample. This indicates that there is room for implementing and facilitating the concept of participatory development and testing, which contribute to the demand orientation and acceptance of AI systems.

Strengths and Limitations

A major strength of this review is the sensitive design of the search strategy, which led to the inclusion of a large sample of study designs and publication types, giving an extensive overview of published works on applications of AI systems in nursing care, as well as considering publications from medical- and informatics-databases and conference archives, which to our knowledge is the first of its kind.

The decision to focus on published works limits the results, as our strategy did not include AI systems already in use for which scientific empirical evidence has not been published and which might be directly introduced to clinical practice by developers. Another notable limitation is our decision to use a rather broad definition of nursing care, care recipients, and care settings, which also included independent living of elderly people. As some of the included publications did not define nursing care as the primary application context, and often included nurses or nursing care facilities as possible users among others, publications dealing with borderline examples of AI application scenarios that might or might not be attributed to the domain of nursing depending on the originating context of the publication, such as medical diagnostics, were not included.

As we used the criterion of conducting field experiments or using real-world data to group studies into basic, experimental, or real-world scenarios, studies in the basic or experimental subsample tested applications that may be considered extending well beyond basic research topics when using other criteria to classify studies. We chose this classification primarily to show the extent of the AI solutions applied in real-world practice. It also needs to be noted that we chose, primarily, absolute and relative numbers of studies and categories to map the existing

literature; however, this provides an overview of prominent research topics and points to gaps in the existing literature; however, this does not allow for an assessment of the sophistication of research on AI approaches done in the context of nursing care. Although an assessment of scientific impact, quality of content, originality, and clarity, as it had been done in the field of sensors, signals, and imaging informatics to identify research works that exemplify recent developments [320] was not part of this review, a number of papers have demonstrated useful techniques to improve the generalizability, interpretability, and reproducibility of increasingly sophisticated models. As we only included publications in English, language bias must be noted. The same holds for the decision to limit the publication range to 2005, and to the selection of the searched databases and conferences, which restrict the sensitivity of the search strategy. However, the increase in publications during the last 5 years indicates that our search managed to cover a relevant period of research and development in AI for nursing care.

Comparison With Previous Work

To our knowledge, this is the first review of its kind to systematize a broad literature base on AI in nursing care, and previous relevant work on this topic is scarce. A direct comparison with the application of AI approaches in other related domains, such as medicine or global health, is difficult, owing to, among other things, reported heterogeneity in AI reporting and the lack of a standardized benchmark [321], which was also present in our sample of studies. Even though disciplines such as biomedicine seem to be more active than nursing in terms of publications targeting AI applications [322], overlapping areas and specific aspects to study such as ethical aspects as well as the impact of digital health interventions and the changes and requirements for the professional role can be noted, for example, in the medicine domain [323]. In relation to the nursing domain, Kikuchi [2] reviewed studies on AI technologies in nursing research that focused on clinical outcomes, such as fall prediction, surgery-related injury, nausea, depression, and survival of patients, as well as on managerial themes addressing bed allocation, decision support, communication risks, nurse burnout, nurses' intention to quit, nursing diagnostics, and knowledge acquisition for nurses. Without including publications outside of medical databases, similar to most studies included in our review, the results indicate a focus on the performance capability of AI algorithms compared with standard statistical methods, underlining the assumption of a lack of evaluation studies on existing AI solutions. Buchanan et al [3] conducted a scoping review on emerging trends in AI-powered health technologies and their implications for domains of nursing, such as administration, clinical practice, policy, and research. In contrast to our current review, specific types of AI approaches were not reported. The described emerging trends are parallel to the application scenarios in this review and entail predictive analytics, clinical decision support systems, smart homes, and health care assistant chatbots, but lay a focus on robot applications [3], which were only sparsely included in this review, as most studies on robots included in the screening process did not meet the inclusion criteria for full text review. The potential of AI to enhance nursing practice, as well as the need for nurses to take on the shared responsibility to influence and take part in the way AI is integrated into the health system, are highlighted by the results [3], contributing to the importance of engaging in the discussion from the perspective of clinical practice and nursing science. Congruent with Shillan et al [70], who reviewed the use of ML approaches in the ICU context, a lack of methodological reporting guidelines for AI approaches using health care data or being conducted in a health care or nursing setting impedes the identification of relevant studies, as well as with the evaluation and rating of real-world relevance, confidence in reported findings, and translation into clinical practice.

Conclusions

Implications for Practice

The aim of this paper was, first, to describe which application scenarios for AI systems in nursing practice were considered in the existing literature. Second, we aimed to show the kinds of AI approaches that have been researched, or are being discussed in the literature and the kinds of care settings involved. Third, we investigated the requirements for the application of AI in nursing practice. Finally, we investigated the ethical, legal, and social aspects of AI and nursing, which are being discussed in the national and international literature.

The results show a broad spectrum of possible application scenarios and facilitate the participation and piloting of existing AI solutions. Because empirical evidence generated in real-world settings is limited, more knowledge on the benefits and advantages of AI approaches, compared with alternative solutions or usual care, is of great need. To date, little is known about the perspectives and experiences of nurses, care dependents, and informal caregivers, who should seek to take an active role in the scientific and societal discourse on AI in nursing care. By educating themselves on the potential harms and benefits of AI applications, they can empower themselves to influence how AI systems will be integrated into their daily lives and practices. Care facilities can contribute to AI development and research by promoting digitalization and ensuring data quality and availability, as successful research and application of AI depends on access, quality, and quantity of data.

Implications for Research

Our results provide an overview of application scenarios for which empirical evidence on algorithm accuracy has been published within the last 15 years. Considering the lack of findings on the effectiveness and application of AI approaches in real-world scenarios, future research should reflect on a more nursing care-specific perspective in their objectives, outcomes, and potential benefits. Aside from clinical, organizational, and managerial outcomes, which can be operationalized from care facility perspectives, our results provide new insights for research activities. Furthermore, discourse on the ethical, legal, and societal implications of AI applications in nursing care, as well as on the participation of stakeholders, needs to be advanced.

Implications for Policy Makers

Half of the publications in our sample have been published during the last 4 years, indicating an increase in research and funding, specifically concerning the application of AI systems in nursing care, with a large uptick of published experimental research. Policy makers and funding bodies might want to reflect on particular priorities for their future grants and programs, against the background of limited empirical evidence of effectiveness and longitudinal evaluation of AI systems. Advanced dissemination of nursing practice with AI technologies also calls for modified qualification, education, and informing of nurses, care dependents, and caregivers. Basic knowledge of AI abilities, opportunities, and limitations, as well as limitations concerning data and AI-generated predictions, could become the subject matter for basic nursing education, as well as practice guidelines and information campaigns to enable nurses to take on a mediating role when implementing AI systems in nursing practice.

Acknowledgments

KWO, DF, and FB conceived the superordinate research project and applied for funding. KS and DD conceptualized this rapid review and performed searches, screening, data extraction, coding, and analyses. KS acted as a second coder for the type of artificial intelligence (AI) system and the application context, as well as wrote the initial manuscript. DB performed data extraction, was the first coder for the type of AI approach and application context, and contributed greatly to the results section of the initial manuscript. DD was the second coder for the type of AI approach and application context. DF, FB, and MSA supervised the development of the search strategy and coding criteria. MSA and FB were the third coders for the type of AI approach and application context in case of disagreement between the first and second coders. All authors commented on and approved the final manuscript. This rapid review is part of the research project “Exploratory project on AI in Nursing” and is funded by the German Ministry for Education and Research (grant #16SV8508).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 33 KB - jmir_v23i11e26522_app1.docx]

Multimedia Appendix 2
Overview of included studies.
[DOCX File , 328 KB - jmir_v23i11e26522_app2.docx]

Multimedia Appendix 3
Overview of studies with applications in real-world settings.
[DOCX File , 77 KB - jmir_v23i11e26522_app3.docx]

References

1. Krick T, Huter K, Domhoff D, Schmidt A, Rothgang H, Wolf-Ostermann K. Digital technology and nursing care: a scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies. BMC Health Serv Res 2019 Jun 20;19(1):400 [FREE Full text] [doi: 10.1186/s12913-019-4238-3] [Medline: 31221133]
2. Kikuchi R. Application of artificial intelligence technology in nursing studies: a systematic review. Online Journal of Nursing Informatics. 2020. URL: https://www.himss.org/resources/application-artificial-intelligence-technology-nursing-studies-systematic-review [accessed 2021-11-17]
3. Buchanan C, Howitt ML, Wilson R, Booth RG, Risling T, Bamford M. Predicted influences of artificial intelligence on the domains of nursing: scoping review. JMIR Nurs 2020 Dec 17;3(1):e23939 [FREE Full text] [doi: 10.2196/23939] [Medline: 34406963]
4. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018 Aug 29;3(4):e000798. [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
5. Jain V, Chatterjee JM. Machine Learning with Health Care Perspective. Cham: Springer; 2020:1-415.
6. Grosan C, Abraham A. Rule-based expert systems. In: Intelligent Systems Intelligent Systems Reference Library. Berlin, Heidelberg: Springer; 2011:149-185.
7. Frize M, Yang L, Walker R, O'Connor A. Conceptual framework of knowledge management for ethical decision-making support in neonatal intensive care. IEEE Trans Inf Technol Biomed 2005 Jun;9(2):205-215. [doi: 10.1109/titb.2005.847187] [Medline: 16138537]

8. Vairaktarakis C, Tsiamis V, Soursou G, Lekkas F, Nikolopoulou M, Vasileiadou E, et al. A computer-aided diagnosis system for geriatrics assessment and frailty evaluation. In: Vlamos P, Alexiou A, editors. GeNeDis 2014. CHA: Springer; 2015:69-77.
9. Lu D, Street WN, Delaney C. Knowledge discovery: detecting elderly patients with impaired mobility. Stud Health Technol Inform 2006;122:121-123. [Medline: 17102231]
10. Zhou F, Jiao JR, Chen S, Zhang D. A case-driven ambient intelligence system for elderly in-home assistance applications. IEEE Trans Syst Man Cybern 2011 Mar;41(2):179-189. [doi: 10.1109/tsmcc.2010.2052456]
11. Tseng KC, Hsu C, Chuang Y. Designing an intelligent health monitoring system and exploring user acceptance for the elderly. J Med Syst 2013 Dec;37(6):9967. [doi: 10.1007/s10916-013-9967-y] [Medline: 24037138]
12. Fodeh SJ, Finch D, Bouayad L, Luther SL, Ling H, Kerns RD, et al. Classifying clinical notes with pain assessment using machine learning. Med Biol Eng Comput 2018 Jul 26;56(7):1285-1292 [FREE Full text] [doi: 10.1007/s11517-017-1772-1] [Medline: 29280092]
13. Tawfik DS, Profit J, Lake ET, Liu JB, Sanders LM, Phibbs CS. Development and use of an adjusted nurse staffing metric in the neonatal intensive care unit. Health Serv Res 2020 Apr 23;55(2):190-200 [FREE Full text] [doi: 10.1111/1475-6773.13249] [Medline: 31869865]
14. Cheng M, Kanai-Pak M, Kuwahara N, Ozaku H, Kogure K, Ota J. Dynamic scheduling based inpatient nursing support: applicability evaluation by laboratory experiments. In: Proceedings of the 3rd ACM International Workshop on Context-Awareness for Self-Managing Systems.: Association for Computing Machinery. 2009; 2009 Presented at: CASEMANS '09: The 3rd ACM International Workshop on Context-Awareness for Self-Managing Systems; 11 May 2009; Nara Japan p. 48-54. [doi: 10.1145/1538864.1538873]
15. Demmer J, Kitzig A, Naroska E. Improvements of a retrospective analysis method for a HMM based posture recognition system in a functionalized nursing bed. In: Proceedings of the IEEE/SICE International Symposium on System Integration (SII). 2017 Presented at: IEEE/SICE International Symposium on System Integration (SII); Dec. 11-14, 2017; Taipei, Taiwan. [doi: 10.1109/SII.2017.8279268]
16. Gerka A, Pfingsthorn M, Lupkes C, Sparenberg K, Frenken M, Lins C, et al. Detecting the number of persons in the bed area to enhance the safety of artificially ventilated persons. In: Proceedings of the IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). 2018 Presented at: IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom); Sept. 17-20, 2018; Ostrava, Czech Republic. [doi: 10.1109/healthcom.2018.8531174]
17. Chen Y, Hsu J, Hung C, Wu J, Lai F, Kuo S. Surgical wounds assessment system for self-care. IEEE Trans Syst Man Cybern Syst 2020 Dec;50(12):5076-5091. [doi: 10.1109/tsmc.2018.2856405]
18. Cho I, Park I, Kim E, Lee E, Bates DW. Using EHR data to predict hospital-acquired pressure ulcers: a prospective study of a Bayesian Network model. Int J Med Inform 2013 Nov;82(11):1059-1067. [doi: 10.1016/j.ijmedinf.2013.06.012] [Medline: 23891086]
19. Marschollek M, Wolf K, Gietzelt M, Nemitz G, Schwabedissen H, Haux R. Assessing elderly persons' fall risk using spectral analysis on accelerometric data--a clinical evaluation study. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:3682-3685. [doi: 10.1109/IEMBS.2008.4650008] [Medline: 19163511]
20. Lynn LA. Artificial intelligence systems for complex decision-making in acute care medicine: a review. Patient Saf Surg 2019;13:6 [FREE Full text] [doi: 10.1186/s13037-019-0188-2] [Medline: 30733829]
21. Peirce AG, Elie S, George A, Gold M, O'Hara K, Rose-Facey W. Knowledge development, technology and questions of nursing ethics. Nurs Ethics 2020 Feb 28;27(1):77-87. [doi: 10.1177/0969733019840752] [Medline: 31032700]
22. Pflegestatistik - Pflege im Rahmen der Pflegeversicherung - Deutschlandergebnisse - 2017. Destatis. 2018. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Pflege/Publikationen/Downloads-Pflege/pflege-deutschlandergebnisse-5224001179004.pdf?__blob=publicationFile, [accessed 2021-11-18]
23. Morley J, Floridi L. An ethically mindful approach to AI for health care. Lancet 2020 Jan 25;395(10220):254-255. [doi: 10.1016/S0140-6736(19)32975-7] [Medline: 31982053]
24. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
25. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019 Sep 02;1(9):389-399. [doi: 10.1038/s42256-019-0088-2]
26. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015 Sep 16;13:224 [FREE Full text] [doi: 10.1186/s12916-015-0465-6] [Medline: 26377409]
27. Seibert K, Domhoff D, Biessmann F, Fürstenau D, Wolf-Ostermann K. A Rapid Review on Application Scenarios for Artificial Intelligence in Nursing Care — Review Protocol. Staats- und Universitätsbibliotheken Bremen. 2020. URL: https://media.suub.uni-bremen.de/handle/elib/4295
28. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

29. Boren SA, Moxley D. Systematically reviewing the literature: building the evidence for health care quality. Mo Med 2015;112(1):58-62 [FREE Full text] [Medline: 25812277]
30. 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]
31. Ackley B, Swan B, Ladwig G, Tucker S. Evidence-Based Nursing Care Guidelines: Medical-surgical Interventions. St Louis, MO: Mosby Elsevier; 2008.
32. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx 2004 Jul;1(3):341-347 [FREE Full text] [doi: 10.1602/neurorx.1.3.341] [Medline: 15717036]
33. Taddy M. The technological elements of artificial intelligence. NBER Working Paper. 2018. URL: https://www.nber.org/system/files/working_papers/w24301/w24301.pdf [accessed 2021-11-18]
34. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620 [FREE Full text] [doi: 10.2196/17620] [Medline: 32406857]
35. Amato F, Bianchi S, Comai S, Crovari P, Pasquarelli M, Imtiaz A, et al. CLONE: a promising system for the remote monitoring of Alzheimer's patients: an experimentation with a wearable device in a village for Alzheimer's care. In: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good.: Association for Computing Machinery; 2018 Presented at: Goodtechs '18: International Conference on Smart Objects and Technologies for Social Good; November 28 - 30, 2018; Bologna Italy p. 255-260. [doi: 10.1145/3284869.3284906]
36. Viswanathan P, Little JJ, Mackworth A, Mihailidis A. An intelligent powered wheelchair for users with dementia: case studies with NOAH (navigation and obstacle avoidance help). Association for the Advancement of Artificial Intelligence 2012:1-6 [FREE Full text]
37. Ye C, Li J, Hao S, Liu M, Jin H, Zheng L, et al. Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. Int J Med Inform 2020 May;137:104105 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104105] [Medline: 32193089]
38. Zampieri FG, Salluh JI, Azevedo LC, Kahn JM, Damiani LP, Borges LP, ORCHESTRA Study Investigators. ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis. Intensive Care Med 2019 Nov 08;45(11):1599-1607. [doi: 10.1007/s00134-019-05790-z] [Medline: 31595349]
39. Zhang Y. Real-time development of patient-specific alarm algorithms for critical care. Annu Int Conf IEEE Eng Med Biol Soc 2007;2007:4351-4354. [doi: 10.1109/IEMBS.2007.4353300] [Medline: 18002966]
40. Davoudi A, Malhotra KR, Shickel B, Siegel S, Williams S, Ruppert M, et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. Sci Rep 2019 May 29;9(1):8020 [FREE Full text] [doi: 10.1038/s41598-019-44004-w] [Medline: 31142754]
41. Carros F, Meurer J, Löffler D, Unbehaun D, Matthies S, Koch I, et al. Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. United States: Association for Computing Machinery; 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-12. [doi: 10.1145/3313831.3376402]
42. Chen D, Bharucha AJ, Wactlar HD. Intelligent video monitoring to improve safety of older persons. Annu Int Conf IEEE Eng Med Biol Soc 2007;2007:3814-3817. [doi: 10.1109/IEMBS.2007.4353163] [Medline: 18002829]
43. Tang V, Siu PK, Choy KL, Lam HY, Ho GT, Lee CK, et al. An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry. Expert Syst 2019 Jan 06;36(2):e12369. [doi: 10.1111/exsy.12369]
44. Xiong GL, Bayen E, Nickels S, Subramaniam R, Agrawal P, Jacquemot J, et al. Real-time video detection of falls in dementia care facility and reduced emergency care. Am J Manag Care 2019 Jul;25(7):314-315 [FREE Full text] [Medline: 31318502]
45. Mervin MC, Moyle W, Jones C, Murfield J, Draper B, Beattie E, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. J Am Med Dir Assoc 2018 Jul;19(7):619-622. [doi: 10.1016/j.jamda.2017.10.008] [Medline: 29325922]
46. Ala-Kitula A, Talvitie-Lamberg K, Tyrväinen P, Silvennoinen M. Developing solutions for healthcare? Deploying artificial intelligence to an evolving target. In: Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI). 2017 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); Dec. 14-16, 2017; Las Vegas, NV, USA. [doi: 10.1109/csci.2017.285]
47. Alwan M. Passive in-home health and wellness monitoring: overview, value and examples. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Sept. 3-6, 2009; Minneapolis, MN, USA. [doi: 10.1109/iembs.2009.5333799]
48. Hirdes JP, Poss JW, Curtin-Telegdi N. The Method for Assigning Priority Levels (MAPLe): a new decision-support system for allocating home care resources. BMC Med 2008 Mar 26;6:9 [FREE Full text] [doi: 10.1186/1741-7015-6-9] [Medline: 18366782]
49. Yang Y, Hirdes JP, Dubin JA, Lee J. Fall risk classification in community-dwelling older adults using a smart wrist-worn device and the resident assessment instrument-home care: prospective observational study. JMIR Aging 2019 Jun 07;2(1):e12153 [FREE Full text] [doi: 10.2196/12153] [Medline: 31518278]

50. Zhu M, Chen W, Hirdes JP, Stolee P. The K-nearest neighbor algorithm predicted rehabilitation potential better than current Clinical Assessment Protocol. J Clin Epidemiol 2007 Oct;60(10):1015-1021. [doi: 10.1016/j.jclinepi.2007.06.001] [Medline: 17884595]
51. Zhu M, Zhang Z, Hirdes JP, Stolee P. Using machine learning algorithms to guide rehabilitation planning for home care clients. BMC Med Inform Decis Mak 2007 Dec 20;7(1):41 [FREE Full text] [doi: 10.1186/1472-6947-7-41] [Medline: 18096079]
52. Bajo J, Corchardo J, Rodriguez S. GR-MAS: multi-agent system for geriatric residences. In: Proceedings of the ECAI 18th European Conference on Artificial Intelligence. 2008 Presented at: ECAI 18th European Conference on Artificial Intelligence; 2008; Patras, Greece.
53. Yamamoto K, Yoshii M, Kinoshita F, Touyama H. Classification vs regression by cnn for handwashing skills evaluations in nursing education. In: Proceedings of the International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2020 Presented at: International Conference on Artificial Intelligence in Information and Communication (ICAIIC); Feb. 19-21, 2020; Fukuoka, Japan. [doi: 10.1109/icaiic48513.2020.9064974]
54. Matsuyama Y, Taniyama H, Fujie S, Kobayashi T. System design of group communication activator: an entertainment task for elderly care. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction. 2009 Presented at: HRI09: International Conference on Human Robot Interaction; March 9 - 13, 2009; La Jolla California USA p. 243-244. [doi: 10.1145/1514095.1514157]
55. Papageorgiou X, Chalvatzaki G, Dometios A, Tzafestas C, Maragos P. Intelligent assistive robotic systems for the elderly: two real-life use cases. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments. 2017 Presented at: PETRA '17: 10th International Conference on PErvasive Technologies Related to Assistive Environments; June 21 - 23, 2017; Island of Rhodes Greece p. 360-365. [doi: 10.1145/3056540.3076184]
56. Tapia D, Rodriguez S, Corchado J. A distributed ambient intelligence based multi-agent system for Alzheimer health care. In: Hassanien AE, Abawajy J, Abraham A, Hagras H, editors. Pervasive Computing. London: Springer; 2009:181-199.
57. Evans R, Kuttler K, Simpson K, Howe S, Crossno P, Johnson K, et al. Automated detection of physiologic deterioration in hospitalized patients. J Am Med Inform Assoc 2015 Mar;22(2):350-360 [FREE Full text] [doi: 10.1136/amiajnl-2014-002816] [Medline: 25164256]
58. Ambagtsheer RC, Shafiabady N, Dent E, Seiboth C, Beilby J. The application of artificial intelligence (AI) techniques to identify frailty within a residential aged care administrative data set. Int J Med Inform 2020 Apr;136:104094. [doi: 10.1016/j.ijmedinf.2020.104094] [Medline: 32058264]
59. Al-Shaqi R, Mourshed M, Rezgui Y. Progress in ambient assisted systems for independent living by the elderly. Springerplus 2016 May 14;5(1):624 [FREE Full text] [doi: 10.1186/s40064-016-2272-8] [Medline: 27330890]
60. Khan SS, Ye B, Taati B, Mihailidis A. Detecting agitation and aggression in people with dementia using sensors - a systematic review. Alzheimers Dement 2018 Jun 20;14(6):824-832. [doi: 10.1016/j.jalz.2018.02.004] [Medline: 29571749]
61. Krishnan RH, Pugazhenthi S. Mobility assistive devices and self-transfer robotic systems for elderly, a review. Intel Serv Robotics 2013 Dec 22;7(1):37-49. [doi: 10.1007/s11370-013-0142-6]
62. Zahia S, Zapirain MB, Sevillano X, González A, Kim PJ, Elmaghraby A. Pressure injury image analysis with machine learning techniques: a systematic review on previous and possible future methods. Artif Intell Med 2020 Jan;102:101742. [doi: 10.1016/j.artmed.2019.101742] [Medline: 31980110]
63. Wai A, Fook F, Jayachandran M, Biswas J, Lee JE, Yap P. Implementation of context-aware distributed sensor network system for managing incontinence among patients with dementia. In: Proceedings of the International Conference on Body Sensor Networks. 2010 Presented at: International Conference on Body Sensor Networks; June 7-9, 2010; Singapore. [doi: 10.1109/bsn.2010.15]
64. Gerke S, Yeung S, Cohen IG. Ethical and legal aspects of ambient intelligence in hospitals. J Am Med Assoc 2020 Feb 18;323(7):601-602. [doi: 10.1001/jama.2019.21699] [Medline: 31977033]
65. Portacolone E, Halpern J, Luxenberg J, Harrison KL, Covinsky KE. Ethical issues raised by the introduction of artificial companions to older adults with cognitive impairment: a call for interdisciplinary collaborations. J Alzheimer's Dis 2020 Jul 21;76(2):445-455. [doi: 10.3233/jad-190952]
66. McCradden MD, Baba A, Saha A, Ahmad S, Boparai K, Fadaiefard P, et al. Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. CMAJ Open 2020 Feb 18;8(1):90-95 [FREE Full text] [doi: 10.9778/cmajo.20190151] [Medline: 32071143]
67. Wangmo T, Lipps M, Kressig RW, Ienca M. Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. BMC Med Ethics 2019 Dec 19;20(1):98 [FREE Full text] [doi: 10.1186/s12910-019-0437-z] [Medline: 31856798]
68. Cesta A, Pecora F, Bickmore T. The ROBOCARE Project: intelligent systems for elder care. AAAI Press. 2005. URL: https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-02/FS05-02-005.pdf [accessed 2021-11-18]
69. Gonzalez-Usach R, Collado V, Esteve M, Palau C. AAL open source system for the monitoring and intelligent control of nursing homes. In: Proceedings of the IEEE 14th International Conference on Networking, Sensing and Control (ICNSC).

2017 Presented at: IEEE 14th International Conference on Networking, Sensing and Control (ICNSC); May 16-18, 2017; Calabria, Italy. [doi: 10.1109/icnsc.2017.8000072]
70. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. Crit Care 2019 Aug 22;23(1):284 [FREE Full text] [doi: 10.1186/s13054-019-2564-9] [Medline: 31439010]
71. Antink CH, Leonhardt S, Walter M. Reducing false alarms in the ICU by quantifying self-similarity of multimodal biosignals. Physiol Meas 2016 Aug 25;37(8):1233-1252. [doi: 10.1088/0967-3334/37/8/1233] [Medline: 27454256]
72. Yu C, Liu J, Zhao H. Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Med Inform Decis Mak 2019 Apr 09;19(Suppl 2):57 [FREE Full text] [doi: 10.1186/s12911-019-0763-6] [Medline: 30961594]
73. Abe A, Ozaku H, Kuwahara N, Kogure K. Cooperation between abductive and inductive nursing risk management. In: Proceedings of the Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06). 2006 Presented at: Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06); Dec. 18-22, 2006; Hong Kong, China. [doi: 10.1109/icdmw.2006.53]
74. Adlassnig K, Blacky A, Koller W. Artificial-intelligence-based hospital-acquired infection control. Stud Health Technol Inform 2009;149:103-110. [Medline: 19745475]
75. Alam M, Heching A, Palmarini N. Scaling longitudinal functional health assessment in multi-inhabitant smarthome. In: Proceedings of the IEEE 39th International Conference on Distributed Computing Systems (ICDCS). 2019 Presented at: IEEE 39th International Conference on Distributed Computing Systems (ICDCS); July 7-10, 2019; Dallas, TX, USA. [doi: 10.1109/icdcs.2019.00217]
76. Alam M, Roy N, Holmes S, Gangopadhyay A, Galik E. Automated functional and behavioral health assessment of older adults with dementia. In: Proceedings of the IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2016 Presented at: IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Washington, DC, USA. [doi: 10.1109/chase.2016.16]
77. Albina E, Hernandez A. Designing an intelligent elderly behavior detection system. In: Proceedings of the Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII). 2018 Presented at: Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII); March 22-24, 2018; Chennai, India. [doi: 10.1109/icbsii.2018.8524658]
78. Alnosayan N, Lee E, Alluhaidan A, Chatterjee S, Houston-Feenstra L, Kagoda M. MyHeart: an intelligent mHealth home monitoring system supporting heart failure self-care. In: Proceedings of the IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom). 2014 Presented at: IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom); Oct. 15-18, 2014; Natal, Brazil. [doi: 10.1109/healthcom.2014.7001860]
79. An Y, Khare R, Song I, Hu X. Data exploration and knowledge discovery in a patient wellness tracking (PWT) system at a nurse-managed health services center. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.: Association for Computing Machinery; 2012 Presented at: IHI '12: ACM International Health Informatics Symposium; January 28 - 30, 2012; Miami Florida USA p. 661-666. [doi: 10.1145/2110363.2110437]
80. Annangi P, Frigstad S, Subin S, Torp A, Ramasubramaniam S, Varna S. An automated bladder volume measurement algorithm by pixel classification using random forests. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug. 16-20, 2016; Orlando, FL, USA. [doi: 10.1109/embc.2016.7591633]
81. Barrué C, Annicchiarico R, Cortés U, Matínez-Velasco A, Martín EX, Campana F, et al. The i-Walker: An Intelligent Pedestrian Mobility Aid. Berlin: Springer; 2010:103-123.
82. Antoniadi AM, Galvin M, Heverin M, Hardiman O, Mooney C. Prediction of caregiver burden in amyotrophic lateral sclerosis: a machine learning approach using random forests applied to a cohort study. BMJ Open 2020 Feb 28;10(2):e033109 [FREE Full text] [doi: 10.1136/bmjopen-2019-033109] [Medline: 32114464]
83. Arif D, Ahmad A, Bakar M, Ihtisham M, Winberg S. Cost effective solution for minimization of medical errors and acquisition of vitals by using autonomous nursing robot. In: Proceedings of the 2017 International Conference on Information System and Data Mining. 2017 Presented at: ICISDM '17: 2017 International Conference on Information System and Data Mining; April 1 - 3, 2017; Charleston SC USA p. 134-138. [doi: 10.1145/3077584.3077598]
84. Asif U, Von Cavallar S, Tang J, Harrer S. SSHFD: single shot human fall detection with occluded joints resilience. arXiv. 2020. URL: https://arxiv.org/abs/2004.00797 [accessed 2021-11-18]
85. Augustyniak P. Adaptive architecture for assisted living systems. In: Proceedings of the 6th International Conference on Human System Interactions (HSI). 2013 Presented at: 6th International Conference on Human System Interactions (HSI); June 6-8, 2013; Sopot, Poland. [doi: 10.1109/hsi.2013.6577881]
86. Au-Yeung WM, Sahani AK, Isselbacher EM, Armoundas AA. Reduction of false alarms in the intensive care unit using an optimized machine learning based approach. NPJ Digit Med 2019;2:86 [FREE Full text] [doi: 10.1038/s41746-019-0160-7] [Medline: 31508497]

87. Aziz O, Klenk J, Schwickert L, Chiari L, Becker C, Park EJ, et al. Validation of accuracy of SVM-based fall detection system using real-world fall and non-fall datasets. PLoS One 2017 Jul;12(7):e0180318 [FREE Full text] [doi: 10.1371/journal.pone.0180318] [Medline: 28678808]
88. Aziz O, Musngi M, Park EJ, Mori G, Robinovitch SN. A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials. Med Biol Eng Comput 2017 Jan;55(1):45-55. [doi: 10.1007/s11517-016-1504-y] [Medline: 27106749]
89. Babic F, Jancus A, Melisova K. Customized web-based system for elderly people using elements of artificial intelligence. In: Proceedings of the 2016 Federated Conference on Computer Science and Information Systems. 2016 Presented at: Federated Conference on Computer Science and Information Systems; 2016; Gdansk, Poland p. 277-280. [doi: 10.15439/2016f165]
90. Bakola I, Bellos C, Bibas A, Koutsouris D, Fotiadis D. An Ambient Intelligent monitoring system to improve the independency of the elderly with balance disorders indoors. In: Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain. [doi: 10.1109/bhi.2014.6864333]
91. Batata O, Augusto V, Xie X. Mixed machine learning and agent-based simulation for respite care evaluation. In: Proceedings of the Winter Simulation Conference (WSC).: IEEE Press; 2018 Presented at: Winter Simulation Conference (WSC); Dec. 9-12, 2018; Gothenburg, Sweden. [doi: 10.1109/wsc.2018.8632385]
92. Baxter GD, Monk AF, Tan K, Dear PR, Newell SJ. Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit. Artif Intell Med 2005 Nov;35(3):243-257. [doi: 10.1016/j.artmed.2005.01.004] [Medline: 15994070]
93. Becker M, Werkman E, Anastasopoulos M, Kleinberger T. Approaching ambient intelligent home care systems. In: Proceedings of the Pervasive Health Conference and Workshops. 2006 Presented at: Pervasive Health Conference and Workshops; Nov. 29 - Dec. 1, 2006; Innsbruck, Austria. [doi: 10.1109/pcthealth.2006.361656]
94. Belshaw M, Taati B, Snoek J, Mihailidis A. Towards a single sensor passive solution for automated fall detection. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:1773-1776 [FREE Full text] [doi: 10.1109/IEMBS.2011.6090506] [Medline: 22254671]
95. Bickmore T, Pfeifer L, Jack B. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: Association for Computing Machinery; 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; April 4 - 9, 2009; Boston MA USA p. 1265-1274. [doi: 10.1145/1518701.1518891]
96. Bizjak J, Gradisek A, Gams M. Intelligent assistant for elderly. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Demos. 2018 Presented at: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Demos; July 13-19, 2018; Stockholm, Sweden p. 5808-5810. [doi: 10.24963/ijcai.2018/845]
97. Bleda A, Maestre R, Beteta M, Vidal J. AmICare: ambient intelligent and assistive system for caregivers support. In: Proceedings of the IEEE 16th International Conference on Embedded and Ubiquitous Computing (EUC). 2018 Presented at: IEEE 16th International Conference on Embedded and Ubiquitous Computing (EUC); Oct. 29-31, 2018; Bucharest, Romania. [doi: 10.1109/euc.2018.00017]
98. Bono M, Sumi Y, Nishida T. Towards achieving complex medical engineering to understand conversational dynamics. In: Proceedings of the IEEE/ICME International Conference on Complex Medical Engineering. 2007 May Presented at: IEEE/ICME International Conference on Complex Medical Engineering; May 23-27, 2007; Beijing, China. [doi: 10.1109/iccme.2007.4381779]
99. Bouchard B, Roy P, Bouzouane A, Giroux S, Mihailidis A. An activity recognition model for Alzheimer's patientsxtension of the COACH task guidance system. In: Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence. 2008 Presented at: ECAI 2008: 18th European Conference on Artificial Intelligence; June 27, 2008; Patras, Greece p. 811-812. [doi: 10.5555/1567281.1567482]
100. Brom H, Carthon JM, Ikeaba U, Chittams J. Leveraging electronic health records and machine learning to tailor nursing care for patients at high risk for readmissions. J Nurs Care Qual 2020;35(1):27-33. [doi: 10.1097/NCQ.0000000000000412] [Medline: 31136529]
101. Campos W, Martinez A, Sanchez W, Estrada H, Castro-Sánchez NA, Mujica D. A systematic review of proposals for the social integration of elderly people using ambient intelligence and social networking sites. Cogn Comput 2016 Feb 12;8(3):529-542. [doi: 10.1007/s12559-016-9382-z]
102. Celi LA, Hinske LC, Alterovitz G, Szolovits P. An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study. Crit Care 2008;12(6):R151 [FREE Full text] [doi: 10.1186/cc7140] [Medline: 19046450]
103. Cetin YE, Sarucan A. Nurse scheduling with opposition-based parallel harmony search algorithm. J Intell Syst 2017;28(4):0150. [doi: 10.1515/jisys-2017-0150]
104. Chen Y, Chiang H, Yu C, Chiang C, Liu C, Wang J. An intelligent knowledge-based and customizable home care system framework with ubiquitous patient monitoring and alerting techniques. Sensors (Basel) 2012 Aug 10;12(8):11154-11186 [FREE Full text] [doi: 10.3390/s120811154] [Medline: 23112650]

105. Cheng L, Zhu M, Poss JW, Hirdes JP, Glenny C, Stolee P. Opinion versus practice regarding the use of rehabilitation services in home care: an investigation using machine learning algorithms. BMC Med Inform Decis Mak 2015 Oct 09;15:80 [FREE Full text] [doi: 10.1186/s12911-015-0203-1] [Medline: 26453354]
106. Chouvarda I, Philip NY, Natsiavas P, Kilintzis V, Sobnath D, Kayyali R, et al. WELCOME – innovative integrated care platform using wearable sensing and smart cloud computing for COPD patients with comorbidities. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3180-3183. [doi: 10.1109/EMBC.2014.6944298] [Medline: 25570666]
107. Cismondi F, Fialho A, Lu X, Vieira S, Gray J, Reti S, et al. ANN validation system for ICU neonatal data. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). 2012 Presented at: International Joint Conference on Neural Networks (IJCNN); June 10-15, 2012; Brisbane, QLD, Australia. [doi: 10.1109/ijcnn.2012.6252782]
108. Clark DG, Kapur P, Geldmacher DS, Brockington JC, Harrell L, DeRamus TP, et al. Latent information in fluency lists predicts functional decline in persons at risk for Alzheimer disease. Cortex 2014 Jun;55:202-218 [FREE Full text] [doi: 10.1016/j.cortex.2013.12.013] [Medline: 24556551]
109. Clavelle JT, Sweeney CD, Swartwout E, Lefton C, Guney S. Leveraging technology to sustain extraordinary care: a qualitative analysis of meaningful nurse recognition. J Nurs Adm 2019 Jun;49(6):303-309. [doi: 10.1097/NNA.0000000000000757] [Medline: 31135637]
110. Cramer EM, Seneviratne MG, Sharifi H, Ozturk A, Hernandez-Boussard T. Predicting the incidence of pressure ulcers in the intensive care unit using machine learning. EGEMS (Wash DC) 2019 Sep 05;7(1):49 [FREE Full text] [doi: 10.5334/egems.307] [Medline: 31534981]
111. Cresswell K, Callaghan M, Khan S, Sheikh Z, Mozaffar H, Sheikh A. Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: A systematic review. Health Informatics J 2020 Sep;26(3):2138-2147 [FREE Full text] [doi: 10.1177/1460458219900452] [Medline: 31964204]
112. Crump C, Saxena S, Wilson B, Farrell P, Rafiq A, Silvers CT. Using Bayesian networks and rule-based trending to predict patient status in the intensive care unit. AMIA Annu Symp Proc 2009 Nov 14;2009:124-128 [FREE Full text] [Medline: 20351835]
113. Cufoglu A, Chin J. Towards an understanding classification of well-being for care of older people. In: Proceedings of the IEEE 13th International Conference on Industrial Informatics (INDIN). 2015 Presented at: IEEE 13th International Conference on Industrial Informatics (INDIN); July 22-24, 2015; Cambridge, UK. [doi: 10.1109/indin.2015.7281954]
114. Cyras K, Karamlou A, Lee M, Letsios D, Misener R, Toni F. AI-assisted Schedule Explainer for Nurse Rostering. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. 2020 Presented at: AAMAS '19: International Conference on Autonomous Agents and Multiagent Systems; May 9 - 13, 2020; Auckland New Zealand p. 2101-2103. [doi: 10.5555/3398761.3399089]
115. Di Pietro TL, Doran DM, McArthur G. Supportive decision making at the point of care: refinement of a case-based reasoning application for use in nursing practice. Comput Inform Nurs 2010;28(4):235-240. [doi: 10.1097/NCN.0b013e3181e1e77a] [Medline: 20571376]
116. Ding B, Chang P, Wang P, Li H, Kuo M. Design of an Intelligent Nursing Clinical Pathway and Nursing Order Support System for Traditional Chinese Medicine. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2017:1014-1018.
117. Dingli A, Abela C. A pervasive assistant for nursing and doctoral staff. In: Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008). 2008 Presented at: 18th European Conference on Artificial Intelligence (ECAI 2008); 27 June 2008; Patras, Greece p. 829-830. [doi: 10.5555/1567281.1567491]
118. Durieux BN, Gramling CJ, Manukyan V, Eppstein MJ, Rizzo DM, Ross LM, et al. Identifying in palliative care consultations: a tandem machine-learning and human coding method. J Palliat Med 2018 Dec;21(12):1755-1760. [doi: 10.1089/jpm.2018.0270] [Medline: 30328760]
119. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A Virtual Agent to Support Individuals Living With Physical and Mental Comorbidities: Co-Design and Acceptability Testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
120. Eerikäinen LM, Vanschoren J, Rooijakkers MJ, Vullings R, Aarts RM. Reduction of false arrhythmia alarms using signal selection and machine learning. Physiol Meas 2016 Aug;37(8):1204-1216. [doi: 10.1088/0967-3334/37/8/1204] [Medline: 27454128]
121. Esfahani S, Muthukumar V, Regentova E, Taghva K, Trabia M. Complex Food Recognition using Hyper-Spectral Imagery. In: Proceedings of the 10th Annual Computing and Communication Workshop and Conference (CCWC). 2020 Presented at: 10th Annual Computing and Communication Workshop and Conference (CCWC); Jan. 6-8, 2020; Las Vegas, NV, USA. [doi: 10.1109/ccwc47524.2020.9031258]
122. Fan X, Zhao Y, Wang H, Tsui KL. Forecasting one-day-forward wellness conditions for community-dwelling elderly with single lead short electrocardiogram signals. BMC Med Inform Decis Mak 2019 Dec 30;19(1):285 [FREE Full text] [doi: 10.1186/s12911-019-1012-8] [Medline: 31888608]
123. Feng P, Yu M, Naqvi S, Chambers J. Deep learning for posture analysis in fall detection. In: Procedings of the 19th International Conference on Digital Signal Processing. 2014 Presented at: 19th International Conference on Digital Signal Processing; Aug. 20-23, 2014; Hong Kong, China. [doi: 10.1109/icdsp.2014.6900806]

124. Feng T, Narayanan S. Modeling Behavioral Consistency in Large-Scale Wearable Recordings of Human Bio-Behavioral Signals. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-8, 2020; Barcelona, Spain. [doi: 10.1109/icassp40776.2020.9054493]
125. Fernandes CO, Miles S, Lucena CJPD, Cowan D. Artificial Intelligence Technologies for Coping with Alarm Fatigue in Hospital Environments Because of Sensory Overload: Algorithm Development and Validation. J Med Internet Res 2019 Nov 26;21(11):e15406 [FREE Full text] [doi: 10.2196/15406] [Medline: 31769762]
126. Fleury A, Vacher M, Noury N. SVM-based multimodal classification of activities of daily living in Health Smart Homes: sensors, algorithms, and first experimental results. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):274-283. [doi: 10.1109/TITB.2009.2037317] [Medline: 20007037]
127. Freed M, Burns B, Heller A, Sanchez D, Beaumont-Bowman S. A Virtual Assistant to Help Dysphagia Patients Eat Safely at Home. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). 2016 Presented at: Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16); July 9–15, 2016; New York USA URL: https://www.ijcai.org/Proceedings/16/Papers/642.pdf
128. Ganesh D, Seshadri G, Sokkanarayanan S, Rajan S, Sathiyanarayanan M. IoT-based Google Duplex Artificial Intelligence Solution for Elderly Care. In: International Conference on Contemporary Computing and Informatics (IC3I). 2019 Presented at: International Conference on contemporary Computing and Informatics (IC3I); Dec. 12-14, 2019; Singapore. [doi: 10.1109/ic3i46837.2019.9055551]
129. Gannod G, Abbott K, Haitsma K, Martindale N, Jennings R, Long C. Using machine learning to facilitate the delivery of person centered care in nursing homes. In: Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018. 2018 Presented at: Thirty-First International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018; May 21-23, 2018; Florida, USA URL: https://www.semanticscholar.org/paper/Using-Machine-Learning-to-Facilitate-the-Delivery-Gannod-Abbott/c363ede498cbb4cc110b78bd4d0bed86b7c1561b#extracted
130. Gargees R, Keller J, Popescu M. Early illness recognition in older adults using transfer learning. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov. 13-16, 2017; Kansas City, MO, USA. [doi: 10.1109/bibm.2017.8217795]
131. Ghose S, Mitra J, Karunanithi M, Dowling J. Human Activity Recognition from Smart-Phone Sensor Data using a Multi-Class Ensemble Learning in Home Monitoring. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2015:62-67.
132. Ghosh N, Maity S, Maity K, Saha S. Non-Parametric Learning Technique for Activity Recognition in Elderly Patients. In: Proceedings of the IEEE Region 10 Conference (TENCON). 2019 Presented at: IEEE Region 10 Conference (TENCON); Oct. 17-20, 2019; Kochi, India. [doi: 10.1109/tencon.2019.8929511]
133. Gochoo M, Tan T, Velusamy V, Liu S, Bayanduuren D, Huang S. Device-Free Non-Privacy Invasive Classification of Elderly Travel Patterns in A Smart House Using PIR Sensors and DCNN. IEEE Sensors J 2020:1-1. [doi: 10.1109/jsen.2017.2771287]
134. Gomes DC, Oliveira LESE, Cubas MR, Barra CMCM. Use of computational tools as support to the cross-mapping method between clinical terminologies. Texto contexto - enferm 2019;28:-. [doi: 10.1590/1980-265x-tce-2017-0187]
135. Gómez-Sebastià I, Moreno J, Álvarez-Napagao S, Garcia-Gasulla D, Barrué C, Cortés U. Situated Agents and Humans in Social Interaction for Elderly Healthcare: From Coaalas to AVICENA. J Med Syst 2016 Feb;40(2):38. [doi: 10.1007/s10916-015-0371-7] [Medline: 26590976]
136. Grgurić A, Mošmondor M, Huljenić D. The SmartHabits: An Intelligent Privacy-Aware Home Care Assistance System. Sensors (Basel) 2019 Feb 21;19(4):- [FREE Full text] [doi: 10.3390/s19040907] [Medline: 30795587]
137. Haigh KZ, Kiff LM, Ho G. The Independent LifeStyle Assistant: lessons learned. Assist Technol 2006 Mar 31;18(1):87-106. [doi: 10.1080/10400435.2006.10131909] [Medline: 16796244]
138. Hever G, Cohen L, O'Connor MF, Matot I, Lerner B, Bitan Y. Machine learning applied to multi-sensor information to reduce false alarm rate in the ICU. J Clin Monit Comput 2020 Apr;34(2):339-352. [doi: 10.1007/s10877-019-00307-x] [Medline: 30955160]
139. Hiissa M, Pahikkala T, Suominen H, Lehtikunnas T, Back B, Karsten H, et al. Towards automated classification of intensive care nursing narratives. Int J Med Inform 2007 Dec;76 Suppl 3:362-368. [doi: 10.1016/j.ijmedinf.2007.03.003] [Medline: 17513166]
140. Howcroft J, Kofman J, Lemaire ED. Feature selection for elderly faller classification based on wearable sensors. J Neuroeng Rehabil 2017 May 30;14(1):47 [FREE Full text] [doi: 10.1186/s12984-017-0255-9] [Medline: 28558724]
141. Hsu C, Chien Y, Yu G, Koppen M, Chen S, Niu X. An intelligent fuzzy affective computing system for elderly living alone. In: Proceedings of the Ninth International Conference on Hybrid Intelligent Systems. 2009 Presented at: Ninth International Conference on Hybrid Intelligent Systems; Aug 12-14, 2009; Shenyang, China. [doi: 10.1109/his.2009.318]

142. Hsu C, Wang M, Shen H, Chiang R, Wen C. FallCare+: An IoT surveillance system for fall detection. In: Proceedings of the 2017 International Conference on Applied System Innovation (ICASI). 2017 Presented at: 2017 International Conference on Applied System Innovation (ICASI); May 13-17, 2017; Sapporo, Japan p. 13-17. [doi: 10.1109/icasi.2017.7988590]
143. Hsu H, Cheng Z, Shih T, Chen C, editors. RFID-based personalized behavior modeling. In: Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing. 2009 Presented at: Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing; Jul 7-9, 2009; Brisbane, QLD, Australia. [doi: 10.1109/uic-atc.2009.29]
144. Hu X, Chu D, Li Z, Dai R, Cui Y, Zhou Z, et al. Coarse-to-fine activity annotation and recognition algorithm for solitary older adults. IEEE Access 2020;8:4051-4064. [doi: 10.1109/access.2019.2962843]
145. Huimin Q, Yaobin M, Wenbo X, Zhiquan W. Home environment fall detection system based on a cascaded multi-SVM classifier. In: Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision. 2008 Presented at: 10th International Conference on Control, Automation, Robotics and Vision; Dec 17-20, 2008; Hanoi, Vietnam p. 17-20. [doi: 10.1109/icarcv.2008.4795758]
146. Hunter J, Gatt A, Portet F, Reiter E, Sripada S. Using natural language generation technology to improve information flows in intensive care units. In: Proceedings of the ECAI 2008 - 18th European Conference on Artificial Intelligence. 2008 Presented at: ECAI 2008 - 18th European Conference on Artificial Intelligence; Jul 21-25, 2008; Patras, Greece. [doi: 10.3233/978-1-58603-891-5-678]
147. Inoue S, Mairittha T, Mairittha N, Hossain T. Integrating activity recognition and nursing care records: the system, experiment, and the dataset. In: Proceedings of the Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). 2019 Presented at: Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR); May 30- Jun 2, 2019; Spokane, WA, USA. [doi: 10.1109/iciev.2019.8858584]
148. Jalal A, Kamal S, Kim D. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. Sensors (Basel) 2014 Jul 02;14(7):11735-11759 [FREE Full text] [doi: 10.3390/s140711735] [Medline: 24991942]
149. Jin S. An intelligent multi-sensor surveillance system for elderly care. SmartCR 2012;2(5):296-307. [doi: 10.6029/smartcr.2012.04.006]
150. Jing Y, Eastwood M, Tan B, Konios A, Hamid A, Collinson M. An intelligent well-being monitoring system for residents in extra care homes. In: Proceedings of the IML 2017: International Conference on Internet of Things and Machine Learning. 2017 Presented at: IML 2017: International Conference on Internet of Things and Machine Learning; Oct 17 - 18, 2017; Liverpool United Kingdom. [doi: 10.1145/3109761.3109769]
151. Johnson WG. CI for a DSS to predict nursing coverage in a geriatric care facility. Stud Health Technol Inform 2006;122:86-90. [Medline: 17102223]
152. Jordan D, Rose SE. Multimedia abstract generation of intensive care data: the automation of clinical processes through AI methodologies. World J Surg 2010 Apr;34(4):637-645. [doi: 10.1007/s00268-009-0319-5] [Medline: 20012610]
153. Juang L, Wu M. Abnormal action tracking using robot vision system. In: Proceedings of the International Joint Conference on Awareness Science and Technology & Ubi-Media Computing (iCAST 2013 & UMEDIA 2013)l action tracking using robot vision system. International Joint Conference on Awareness Science and Technology & Ubi-Media Computing (iCAST & UMEDIA). 2013 Presented at: International Joint Conference on Awareness Science and Technology & Ubi-Media Computing (iCAST 2013 & UMEDIA 2013)l action tracking using robot vision system. International Joint Conference on Awareness Science and Technology & Ubi-Media Computing (iCAST & UMEDIA); Nov. 2-4, 2013; Aizu-Wakamatsu, Japan p. 2-4. [doi: 10.1109/icawst.2013.6765480]
154. Juang L, Wu M. Fall down detection under smart home system. J Med Syst 2015 Oct 15;39(10):107. [doi: 10.1007/s10916-015-0286-3] [Medline: 26276014]
155. Jung H, Park H. Development and evaluation of a prototype CDSS for fall prevention. Stud Health Technol Inform 2019 Aug 21;264:1700-1701. [doi: 10.3233/SHTI190604] [Medline: 31438300]
156. Jung K, Covington S, Sen CK, Januszyk M, Kirsner RS, Gurtner GC, et al. Rapid identification of slow healing wounds. Wound Repair Regen 2016;24(1):181-188 [FREE Full text] [doi: 10.1111/wrr.12384] [Medline: 26606167]
157. Kaewprag P, Newton C, Vermillion B, Hyun S, Huang K, Machiraju R. Predictive modeling for pressure ulcers from intensive care unit electronic health records. AMIA Jt Summits Transl Sci Proc 2015;2015:82-86 [FREE Full text] [Medline: 26306245]
158. Kalidas V, Tamil LS. Cardiac arrhythmia classification using multi-modal signal analysis. Physiol Meas 2016 Aug 25;37(8):1253-1272. [doi: 10.1088/0967-3334/37/8/1253] [Medline: 27454417]
159. Kang KJ, Ka B, Kim SJ. A service scenario generation scheme based on association rule mining for elderly surveillance system in a smart home environment. Eng Appl Artif Intell 2012 Oct;25(7):1355-1364. [doi: 10.1016/j.engappai.2012.02.003]
160. Kao H, Hosseinmardi H, Yan S, Hasan M, Narayanan S, Lerman K. Discovering Latent Psychological Structures from Self-Report Assessments of Hospital Workers. In: Proceedings of the 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC). 2018 Presented at: 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC); Nov. 12-14, 2018; Kaohsiung, Taiwan. [doi: 10.1109/besc.2018.8697325]

161. van Kasteren TL, Englebienne G, Kröse BJ. An activity monitoring system for elderly care using generative and discriminative models. Pers Ubiquit Comput 2010 Feb 26;14(6):489-498. [doi: 10.1007/s00779-009-0277-9]
162. Khan S, Zhu T, Ye B, Mihailidis A, Iaboni A, Newman K. DAAD: a framework for detecting agitation and aggression in people living with dementia using a novel multi-modal sensor network. 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, LA, USA p. 18-21. [doi: 10.1109/icdmw.2017.98]
163. Khosla R, Chu M. Embodying care in Matilda. ACM Trans Manage Inf Syst 2013 Dec;4(4):1-33. [doi: 10.1145/2544104]
164. Kingetsu H, Konno T, Awai S, Fukuda D, Sonoda T. Video-based Fall Risk Detection System for the Elderly. In: Proceedings of the IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech). 2019 Presented at: IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech); March 12-14, 2019; Osaka, Japan. [doi: 10.1109/lifetech.2019.8883958]
165. Kobayashi Y, Yamazaki K, Yamazaki A, Gyoda M, Tabata T, Kuno Y. Care robot able to show the order of service provision through bodily actions in multi-party settings. In: Proceedings of the CHI EA '12: CHI '12 Extended Abstracts on Human Factors in Computing Systems. 2012 Presented at: CHI EA '12: CHI '12 Extended Abstracts on Human Factors in Computing Systems; May 5 - 10, 2012; Austin Texas USA. [doi: 10.1145/2212776.2223724]
166. Krüger F, Heine C, Bader S, Hein A, Teipel S, Kirste T. On the applicability of clinical observation tools for human activity annotation. In: Procedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2017 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 13-17, 2017; Kona, HI, USA. [doi: 10.1109/percomw.2017.7917545]
167. Kubota N, Hiroyuki K, Taniguchi K, Sawayama T. State estimation based on a spiking neural network using ultrasonic oscillosensors. In: Proceedings of the World Automation Congress. 2008 Presented at: World Automation Congress; Sept. 28 - Oct. 2, 2008; Waikoloa, HI, USA URL: https://ieeexplore.ieee.org/document/4698994?arnumber=4698994
168. Kumar M, Teso S, Causmaecker P, Raedt L. Automating personnel rostering by learning constraints using tensors. In: Proceedings of the IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). 2019 Presented at: IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI); Nov. 4-6, 2019; Portland, OR, USA. [doi: 10.1109/ictai.2019.00102]
169. Kundu S, Acharyya S. A SAT approach for solving the nurse scheduling problem. In: Proceedings of the TENCON 2008 - IEEE Region 10 Conference. 2008 Presented at: TENCON 2008 - IEEE Region 10 Conference; Nov. 19-21, 2008; Hyderabad, India. [doi: 10.1109/tencon.2008.4766380]
170. Kuspinar A, Hirdes JP, Berg K, McArthur C, Morris JN. Development and validation of an algorithm to assess risk of first-time falling among home care clients. BMC Geriatr 2019 Oct 14;19(1):264 [FREE Full text] [doi: 10.1186/s12877-019-1300-2] [Medline: 31610776]
171. Lea C, Facker J, Hager G, Taylor R, Saria S. 3d sensing algorithms towards building an intelligent intensive care unit. AMIA Jt Summits Transl Sci Proc 2013;2013:136-140 [FREE Full text] [Medline: 24303253]
172. Lee H, Kim J, Yang D, Kim JH. Embedded real-time fall detection using deep learning for elderly care. arXiv. 2017. URL: https://arxiv.org/abs/1711.11200 [accessed 2021-11-18]
173. Lehman EP, Krishnan RG, Zhao X, Mark RG, Lehman LH. Representation learning approaches to detect false arrhythmia alarms from ECG dynamics. Proc Mach Learn Res 2018 Aug;85:571-586 [FREE Full text] [Medline: 31723938]
174. Xiaochun L, jianan C, Guangcai L, Feilong C, Longshen W. A design and implementation of an intelligent care system for the elderly based on internet of things. In: Proceedings of the Eleventh International Conference on Advanced Computational Intelligence (ICACI). 2019 Presented at: Eleventh International Conference on Advanced Computational Intelligence (ICACI); June 7-9, 2019; Guilin, China. [doi: 10.1109/icaci.2019.8778502]
175. Li D, Mathews C. Automated measurement of pressure injury through image processing. J Clin Nurs 2017 Nov;26(21-22):3564-3575. [doi: 10.1111/jocn.13726] [Medline: 28071843]
176. Li M, Xu G, He B, Ma X, Xie J. Pre-impact fall detection based on a modified zero moment point criterion using data from kinect sensors. IEEE Sensors J 2018 Jul 1;18(13):5522-5531. [doi: 10.1109/jsen.2018.2833451]
177. Li Q, Clifford GD. Signal quality and data fusion for false alarm reduction in the intensive care unit. J Electrocardiol 2012;45(6):596-603. [doi: 10.1016/j.jelectrocard.2012.07.015] [Medline: 22960167]
178. Liao P, Hsu P, Chu W, Chu W. Applying artificial intelligence technology to support decision-making in nursing: a case study in Taiwan. Health Informatics J 2015 Jun;21(2):137-148. [doi: 10.1177/1460458213509806] [Medline: 26021669]
179. Lim S, Chung L, Han O, Kim JH. An interactive cyber-physical system (CPS) for people with disability and frail elderly people. In: Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication. 2011 Presented at: ICUIMC '11: The 5th International Conference on Ubiquitous Information Management and Communication; February 21 - 23, 2011; Seoul Korea p. 1-8. [doi: 10.1145/1968613.1968745]
180. Liu W, Chuang K, Chen K. The Design and Implementation of a Chatbot's Character for Elderly Care. In: Proceedings of the International Conference on System Science and Engineering (ICSSE). 2018 Presented at: International Conference on System Science and Engineering (ICSSE); June 28-30, 2018; New Taipei, Taiwan. [doi: 10.1109/icsse.2018.8520008]
181. Lo C, Lin C, Wang C, Dai T, Wong D. Artificial immune systems for intelligent nurse rostering. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management. 2007 Presented at: IEEE International

Conference on Industrial Engineering and Engineering Management; Dec. 2-4, 2007; Singapore. [doi: 10.1109/ieem.2007.4419313]
182. Lodhi MK, Ansari R, Yao Y, Keenan GM, Wilkie D, Khokhar AA. Predicting hospital re-admissions from nursing care data of hospitalized patients. Adv Data Min 2017;2017:181-193 [FREE Full text] [doi: 10.1007/978-3-319-62701-4_14] [Medline: 29104962]
183. Luo L, Li J, Liu C, Shen W. Using machine-learning methods to support health-care professionals in making admission decisions. Int J Health Plann Manage 2019 Apr 07;34(2):1236-1246. [doi: 10.1002/hpm.2769] [Medline: 30957270]
184. Luo S, Zhao C, Fu Y. An intelligent human activity recognition method with incremental learning capability for bedridden patients. In: Proceedings of the 15th International Conference on Control, Automation, Robotics and Vision (ICARCV). 2018 Presented at: 15th International Conference on Control, Automation, Robotics and Vision (ICARCV); Nov. 18-21, 2018; Singapore. [doi: 10.1109/icarcv.2018.8581232]
185. Luperto M, Romeo M, Lunardini F, Basilico N, Jones R, Cangelosi A, et al. Digitalized cognitive assessment mediated by a virtual caregiver. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Demos. 2018 Presented at: Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18); July 13 - 19, 2018; Stockholm Sweden p. 5841-5843. [doi: 10.24963/ijcai.2018/856]
186. Lutrek M, Kalua B, Cvetkovi B, Dovgan E, Gjoreski H, Mirchevska V, et al. Confidence: Ubiquitous care system to support independent living. In: Proceedings of the European Conference on Artificial Intelligence – ECAI. 2012 Presented at: European Conference on Artificial Intelligence – ECAI; Aug 27-31 , 2012; France URL: https://www.researchgate.net/publication/289078287_Confidence_Ubiquitous_care_system_to_support_independent_living
187. Luu T, Phan R, Davey R, Chetty G, editors. A multilevel NER framework for automatic clinical name entity recognition. 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, LA, USA. [doi: 10.1109/icdmw.2017.161]
188. Machanje D, Orero J, Marsala C. Distress recognition from speech analysis: a pairwise association rules-based approach. In: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI). 2019 Presented at: IEEE Symposium Series on Computational Intelligence (SSCI); Dec. 6-9, 2019; Xiamen, China. [doi: 10.1109/ssci44817.2019.9002972]
189. Magyar J, Kobayashi M, Nishio S, Sin?ák P, Ishiguro H. Autonomous robotic dialogue system with reinforcement learning for elderlies with dementia. In: Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). 2019 Presented at: IEEE International Conference on Systems, Man and Cybernetics (SMC); Oct. 6-9, 2019; Bari, Italy. [doi: 10.1109/smc.2019.8914248]
190. Mamun K, Sharma A, Hoque A, Szecsi T. Remote patient physical condition monitoring service module for iWARD hospital robots. In: Proceedings of the Asia-Pacific World Congress on Computer Science and Engineering. 2014 Presented at: Asia-Pacific World Congress on Computer Science and Engineering; Nov. 4-5, 2014; Nadi, Fiji. [doi: 10.1109/apwccse.2014.7053854]
191. Manna T, Swetapadma A, Abdar M. Decision tree predictive learner-based approach for false alarm detection in ICU. J Med Syst 2019 May 21;43(7):191. [doi: 10.1007/s10916-019-1337-y] [Medline: 31115734]
192. Mano L. Emotional condition in the health smart homes environment: emotion recognition using ensemble of classifiers. In: Proceedings of the Innovations in Intelligent Systems and Applications (INISTA). 2018 Presented at: Innovations in Intelligent Systems and Applications (INISTA); July 3-5, 2018; Thessaloniki, Greece. [doi: 10.1109/inista.2018.8466318]
193. Marafino BJ, Boscardin WJ, Dudley RA. Efficient and sparse feature selection for biomedical text classification via the elastic net: application to ICU risk stratification from nursing notes. J Biomed Inform 2015 Apr;54:114-120 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.003] [Medline: 25700665]
194. Masafumi Y, Tomohiro Y, Takeshi F. Study on effect of MOGA with interactive island model using visualization. In: Proceedings of the IEEE Congress on Evolutionary Computation. 2010 Presented at: IEEE Congress on Evolutionary Computation; July 18-23, 2010; Barcelona, Spain. [doi: 10.1109/cec.2010.5585950]
195. Merico D, Bisiani R, Malizia F, Rizzi G, Ali H. Demonstrating contexta-CARE: a situation-aware system for supporting independent living. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; May 5 - Aug 5, 2013; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252048]
196. Mihailidis A, Boger J, Canido M, Hoey J. The use of an intelligent prompting system for people with dementia. Interactions 2007 Jul;14(4):34-37. [doi: 10.1145/1273961.1273982]
197. Mitabe N, Shinomiya N. An intelligent care support system for the elderly with an autonomous robot in ambient sensors. In: Proceedings of the IEEE 6th Global Conference on Consumer Electronics (GCCE). 2017 Presented at: IEEE 6th Global Conference on Consumer Electronics (GCCE); Oct. 24-27, 2017; Nagoya, Japan. [doi: 10.1109/gcce.2017.8229483]
198. Mitabe N, Shinomiya N. An intelligent support system for elderly care with RFID tags and a cleaning robot. In: Proceedings of the 12th International Conference on Sensing Technology (ICST). 2018 Presented at: 12th International Conference on Sensing Technology (ICST); 2018; Limerick, Ireland. [doi: 10.1109/icsenst.2018.8603574]

199. Miyawaki F, Masamune K, Suzuki S, Yoshimitsu K, Vain J. Scrub nurse robot system—intraoperative motion analysis of a scrub nurse and timed-automata-based model for surgery. IEEE Trans Ind Electron 2005 Oct;52(5):1227-1235. [doi: 10.1109/tie.2005.855692]
200. Mori Y, Ikeura R, Ding M. Using tactile sensors to estimate care receiver position on dual arms of robot. In: Proceedings of the IEEE Conference SENSORS. 2013 Presented at: IEEE Conference SENSORS; Nov. 3-6, 2013; Baltimore, MD, USA. [doi: 10.1109/icsens.2013.6688439]
201. Morita K, Fuketa M, Aoe JI, Yasuda K. Improved dialogue communication systems for individuals with dementia. Int J Comput Appl Technol 2015;52(2/3):127. [doi: 10.1504/ijcat.2015.071973]
202. Morita T, Taki K, Fujimoto M, Suwa H, Arakawa Y, Yasumoto K. BLE beacon-based activity monitoring system toward automatic generation of daily report. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2018 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 19-23, 2018; Athens, Greece. [doi: 10.1109/percomw.2018.8480348]
203. Muckell J, Young Y, Leventhal M. A wearable motion tracking system to reduce direct care worker injuries. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH '17: International Conference on Digital Health; July 2 - 5, 2017; London United Kingdom p. 1-6. [doi: 10.1145/3079452.3079493]
204. Naya F, Ohmura R, Miyamae M, Noma H, Kogure K, Imai M. Wireless sensor network system for supporting nursing context-awareness. Int J Autonom Adapt Commun Syst 2011;4(4):361. [doi: 10.1504/ijaacs.2011.043477]
205. Nguyen-Truong CK, Fritz RL. Health-assistive smart homes for aging in place: leading the way for integration of the asian immigrant minority voice. Asian Pac Isl Nurs J 2018;3(4):154-159 [FREE Full text] [doi: 10.31372/20180304.1087] [Medline: 31037263]
206. Nie G, Ullal A, Swanson A, Weitlauf A, Warren Z, Sarkar N. Design of an intelligent and immersive system to facilitate the social interaction between caregivers and young children with autism. In: Proceedings of the International Conference on Human-Computer Interaction. 2019 Presented at: HCII 2019: International Conference on Human-Computer Interaction Universal Access in Human-Computer Interaction Multimodality and Assistive Environments; July 26-31, 2019; Orlando, FL, USA. [doi: 10.1007/978-3-030-23563-5_11]
207. Nii M, Ando S, Takahashi Y, Uchinuno A, Sakashita R. Feature extraction from nursing-care texts for classification. In: World Automation Congress. 2008 Presented at: World Automation Congress; Sept. 28 - Oct. 2, 2008; Waikoloa, HI, USA p. - URL: https://ieeexplore.ieee.org/document/4698973?arnumber=4698973
208. Nii M, Ando S, Takahashi Y, Uchinuno A, Sakashita R. Nursing-care freestyle text classification using support vector machines. In: Proceedings of the IEEE International Conference on Granular Computing (GRC). 2007 Presented at: IEEE International Conference on Granular Computing (GRC); Nov. 2-4, 2007; Fremont, CA, USA. [doi: 10.1109/grc.2007.131]
209. Nii M, Miyake S, Takahama K, Uchinuno A, Sakashita R. Consideration about utilizing text architecture for making feature vectors in classifying nursing-care texts. In: Proceedings of the IEEE International Conference on Systems, Man, Cybernetics. 2013 Presented at: IEEE International Conference on Systems, Man, Cybernetics; Oct. 13-16, 2013; Manchester, UK p. 13-16. [doi: 10.1109/smc.2013.313]
210. Nii M, Okajima S, Sakashita R, Hamada M, Kobashi S. Tongue movement classification in chewing and swallowing using electromyography. In: Proceedings of the 6th International Conference on Informatics, Electronics and Vision &amp 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT). 2017 Presented at: 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT); Sept. 1-3, 2017; Himeji, Japan. [doi: 10.1109/iciev.2017.8338594]
211. Nii M, Takahama K, Uchinuno A, Sakashita R. A directed graph based feature definition for classifying nursing-care texts. In: Proceedings of the IEEE International Conference on Systems, Man, Cybernetics (SMC). 2014 Presented at: IEEE International Conference on Systems, Man,Cybernetics (SMC); Oct. 5-8, 2014; San Diego, CA, USA. [doi: 10.1109/smc.2014.6974504]
212. Nii M, Takahama K, Uchinuno A, Sakashita R. A matrix-based feature vector definition and a SVM-BDT-based classification system for classifying nursing-care texts. In: Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). 2015 Presented at: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); Aug. 2-5, 2015; Istanbul, Turkey. [doi: 10.1109/fuzz-ieee.2015.7337976]
213. Nii M, Takahama K, Uchinuno A, Sakashita R. Soft class decision for nursing-care text classification using a k-nearest neighbor based system. In: Proceedings of the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). 2014 Presented at: 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); July 6-11, 2014; Beijing, China. [doi: 10.1109/fuzz-ieee.2014.6891739]
214. Nii M, Tuchida Y, Iwamoto T, Uchinuno A, Sakashita R. Nursing-care text evaluation using word vector representations realized by word2vec. In: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). 2016 Jul Presented at: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); July 24-29, 2016; Vancouver, BC, Canada. [doi: 10.1109/fuzz-ieee.2016.7737960]

215. Nii M, Yamaguchi T, Takahashi Y, Sakashita R, Uchinuno A. Analysis of nursing-care freestyle Japanese text classification using ga-based term selection. In: Proceedings of the World Automation Congress. 2010 Presented at: World Automation Congress; Sept 19-23, 2010; Kobe, Japan.
216. Nii M, Yamaguchi T, Takahashi Y, Uchinuno A, Sakashita R. Fuzzy Rule Extraction from Nursing-Care Texts. In: Proceedings of the 39th International Symposium on Multiple-Valued Logic. 2009 Presented at: 39th International Symposium on Multiple-Valued Logic; May 21-23, 2009; Naha, Japan. [doi: 10.1109/ismvl.2009.61]
217. Niiuchi R, Kang H, Iwamura K. Detection of human motion gestures using machine learning for actual emergency situations. In: Proceedings of the IEEE 7th Global Conference on Consumer Electronics (GCCE). 2018 Presented at: IEEE 7th Global Conference on Consumer Electronics (GCCE); Oct. 9-12, 2018; Nara, Japan. [doi: 10.1109/gcce.2018.8574609]
218. Nishio S, Hossain B, Yagi N, Nii M, Hiranaka T, Kobashi S. Surgical phase recognition method with a sequential consistency for CAOS-AI navigation system. In: Proceedings of the IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech). 2020 Presented at: IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech); March 10-12, 2020; Kyoto, Japan. [doi: 10.1109/lifetech48969.2020.1570619203]
219. Nouira K, Trabelsi A. Intelligent monitoring system for intensive care units. J Med Syst 2012 Aug;36(4):2309-2318. [doi: 10.1007/s10916-011-9698-x] [Medline: 21505862]
220. Nuutinen M, Leskelä RL, Suojalehto E, Tirronen A, Komssi V. Development and validation of classifiers and variable subsets for predicting nursing home admission. BMC Med Inform Decis Mak 2017 Apr 13;17(1):39 [FREE Full text] [doi: 10.1186/s12911-017-0442-4] [Medline: 28407806]
221. O'Brien K, Liggett A, Ramirez-Zohfeld V, Sunkara P, Lindquist LA. Voice-controlled intelligent personal assistants to support aging in place. J Am Geriatr Soc 2020 Jan 16;68(1):176-179. [doi: 10.1111/jgs.16217] [Medline: 31617581]
222. Ohura N, Mitsuno R, Sakisaka M, Terabe Y, Morishige Y, Uchiyama A, et al. Convolutional neural networks for wound detection: the role of artificial intelligence in wound care. J Wound Care 2019 Oct 01;28(Sup10):13-24. [doi: 10.12968/jowc.2019.28.Sup10.S13] [Medline: 31600101]
223. Okano T, Kitakoshi D, Suzuki M. A preliminary study on preventive care system based on game playing with communication robots. In: Proceedings of the Conference on Technologies and Applications of Artificial Intelligence. 2013 Presented at: 2013 Conference on Technologies and Applications of Artificial Intelligence; Dec. 6-8, 2013; Taipei, Taiwan. [doi: 10.1109/taai.2013.29]
224. Ongenae F, Duysburgh P, Verstraete M, Sulmon N, Bleumers L, Jacobs A. User-driven design of a context-aware application: an ambient-intelligent nurse call system. In: Proceedings of the User-Centered Design of Pervasive Healthcare Applications Conference. 2012 Presented at: User-Centered Design of Pervasive Healthcare Applications Conference; May 21, 2012; San Diego, United States. [doi: 10.4108/icst.pervasivehealth.2012.248699]
225. Ostojic D, Guglielmini S, Moser V, Fauchère JC, Bucher HU, Bassler D, et al. Reducing false alarm rates in neonatal intensive care: a new machine learning approach. Adv Exp Med Biol 2020;1232:285-290. [doi: 10.1007/978-3-030-34461-0_36] [Medline: 31893422]
226. Pahl C, Varadarajan K. Emotional feedback for service robots using hapto-acoustic interface. In: Proceedings of the Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech). 2015 Presented at: Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech); Nov. 26-27, 2015; Port Elizabeth, South Africa. [doi: 10.1109/robomech.2015.7359526]
227. Park C, Kim J, Chen J, Ma H, Stojmenovic I. A location and emergency monitoring system for elder care using ZigBee. In: Proceedings of the Seventh International Conference on Mobile Ad-hoc and Sensor Networks. 2011 Presented at: Seventh International Conference on Mobile Ad-hoc and Sensor Networks; Dec.16-18, 2011; Beijing, China. [doi: 10.1109/msn.2011.87]
228. Park JI, Bliss DZ, Chi C, Delaney CW, Westra BL. Knowledge discovery with machine learning for hospital-acquired catheter-associated urinary tract infections. Comput Inform Nurs 2020 Jan;38(1):28-35 [FREE Full text] [doi: 10.1097/CIN.0000000000000562] [Medline: 31524687]
229. Park S, Won DD, Lee BJ, Escobedo D, Esteva A, Aalipour A, et al. A mountable toilet system for personalized health monitoring via the analysis of excreta. Nat Biomed Eng 2020 Jun;4(6):624-635 [FREE Full text] [doi: 10.1038/s41551-020-0534-9] [Medline: 32251391]
230. Pascual J, Sanz-Bobi M, Contreras D, Azevedo L, Londral A. Intelligent system for assisting elderly people at home. In: Proceedings of the First International Conference on Health Informatics. 2008 Presented at: First International Conference on Health Informatics; Jan 28-31, 2008; Funchal, Madeira, Portugal. [doi: 10.5220/0001036901590163]
231. Patrick J, Li M. An ontology for clinical questions about the contents of patient notes. J Biomed Inform 2012 Apr;45(2):292-306 [FREE Full text] [doi: 10.1016/j.jbi.2011.11.008] [Medline: 22142949]
232. Patterson BW, Engstrom CJ, Sah V, Smith MA, Mendonça EA, Pulia MS, et al. Training and interpreting machine learning algorithms to evaluate fall risk after emergency department visits. Med Care 2019 Jul;57(7):560-566 [FREE Full text] [doi: 10.1097/MLR.0000000000001140] [Medline: 31157707]
233. Paul A, Wan E, Adenwala F, Schafermeyer E, Preiser N, Kaye J, et al. MobileRF: a robust device-free tracking system based on a hybrid neural network HMM classifier. Proc ACM Int Conf Ubiquitous Comput 2014;2014:159-170 [FREE Full text] [doi: 10.1145/2632048.2632097] [Medline: 25544964]

234. Pessemier T, Dooms S, Martens L. A food recommender for patients in a care facility. In: Proceedings of the 7th ACM conference on Recommender systems. 2013 Presented at: RecSys '13: 7th ACM conference on Recommender systems; October 12 - 16, 2013; Hong Kong China. [doi: 10.1145/2507157.2507198]
235. Phan R, Luu T, Davey R, Chetty G. Enhancing clinical name entity recognition based on hybrid deep learning scheme. In: Proceedings of the InternatioCnal onference on Data Mining Workshops (ICDMW). 2019 Presented at: International Conference on Data Mining Workshops (ICDMW); Nov. 8-11, 2019; Beijing, China. [doi: 10.1109/icdmw.2019.00153]
236. Ramanujam E, Padmavathi S. A vision-based posture monitoring system for the elderly using intelligent fall detection technique. In: Guide to Ambient Intelligence in the IoT Environment. Cham: Springer; 2019:249-269.
237. Richter J, Hirtz G. Novel methods for feature extraction based on motion history images and evaluation with regard to altering viewing angles. In: Proceedings of the IEEE Third International Conference on Consumer Electronics ¿ Berlin (ICCE-Berlin)el methods for feature extraction based on motion history images and evaluation with regard to altering viewing angles. IEEE Third International Conference on Consumer Electronics, Berlin (ICCE-Berlin). 2013 Presented at: IEEE Third International Conference on Consumer Electronics ¿ Berlin (ICCE-Berlin)el methods for feature extraction based on motion history images and evaluation with regard to altering viewing angles. IEEE Third International Conference on Consumer Electronics, Berlin (ICCE-Berlin); Sept.9-11, 2013; Berlin, Germany. [doi: 10.1109/icce-berlin.2013.6697978]
238. Robben S, Bosch L, Wiggers P, Decancq J, Kanis M. Managing flexible care with a context aware system for ageing-in-place. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth).: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering); 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); May 20-23, 2015; Istanbul, Turkey. [doi: 10.4108/icst.pervasivehealth.2015.259229]
239. Rouleau G, Gagnon M, Côté J, Payne-Gagnon J, Hudson E, Dubois C. Impact of information and communication technologies on nursing care: results of an overview of systematic reviews. J Med Internet Res 2017 Apr 25;19(4):e122 [FREE Full text] [doi: 10.2196/jmir.6686] [Medline: 28442454]
240. Sadahiro T, Hamazaki M, Miyawaki F, Yoshimitsu K, Masamune K. Laparoscopic skill measurement with COP to realize a HAM Scrub Nurse Robot system. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. 2007 Presented at: IEEE International Conference on Systems, Man and Cybernetics; Oct. 7-10, 2007; Montreal, QC, Canada. [doi: 10.1109/icsmc.2007.4413969]
241. Sadreazami H, Bolic M, Rajan S. On the use of ultra wideband radarstacked LSTM-RNN for at home fall detection. In: Proceedings of the IEEE Life Sciences Conference (LSC). 2018 Presented at: IEEE Life Sciences Conference (LSC); Oct. 28-30, 2018; Montreal, QC. Canada. [doi: 10.1109/lsc.2018.8572048]
242. Sadreazami H, Bolic M, Rajan S. Residual network-based supervised learning of remotely sensed fall incidents using ultra-wideband radar. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS). 2019 Presented at: IEEE International Symposium on Circuits and Systems (ISCAS); May 26-29, 2019; Sapporo, Japan. [doi: 10.1109/iscas.2019.8702446]
243. Sadreazami H, Bolic M, Rajan S. CapsFall: fall detection using ultra-wideband radar and capsule network. IEEE Access 2019;7:55336-55343. [doi: 10.1109/access.2019.2907925]
244. Sakakibara S, Saiki S, Nakamura M, Yasuda K, Duffy V. Generating personalized dialogue towards daily counseling system for home dementia care. In: Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety. Cham: Springer; 2017.
245. Samra H, Li A, Soh B, Alzain M. A conceptual model for an intelligent simulation-based learning management system using a data mining agent in clinical skills education. In: Proceedings of the 4th International Conference on Enterprise Systems (ES). 2016 Presented at: 4th International Conference on Enterprise Systems (ES); Nov. 2-3, 2016; Melbourne, VIC, Australia. [doi: 10.1109/es.2016.17]
246. Scalzo F, Hu X. Semi-supervised detection of intracranial pressure alarms using waveform dynamics. Physiol Meas 2013 Apr;34(4):465-478. [doi: 10.1088/0967-3334/34/4/465] [Medline: 23524637]
247. Scalzo F, Liebeskind D, Hu X. Reducing false intracranial pressure alarms using morphological waveform features. IEEE Trans Biomed Eng 2013 Jan;60(1):235-239 [FREE Full text] [doi: 10.1109/TBME.2012.2210042] [Medline: 22851230]
248. Sharma V, Sharma V, Khan A, Wassmer DJ, Schoenholtz MD, Hontecillas R, et al. Malnutrition, health and the role of machine learning in clinical setting. Front Nutr 2020;7:44 [FREE Full text] [doi: 10.3389/fnut.2020.00044] [Medline: 32351968]
249. Shen V, Horng-Yih L, Ah-Fur L. Application of high-level fuzzy petri nets to fall detection system using smartphone. In: Proceedings of the International Conference on Machine Learning and Cybernetics. 2013 Presented at: International Conference on Machine Learning and Cybernetics; July 14-17, 2013; Tianjin, China. [doi: 10.1109/icmlc.2013.6890807]
250. Shieh Y, Shieh M. Consumer electronics-based intelligent alert system for unattended elderly residents. In: Procedings of the IEEE International Symposium on Consumer Electronics (ISCE). 2013 Presented at: IEEE International Symposium on Consumer Electronics (ISCE); June 3-6, 2013; Hsinchu, Taiwan. [doi: 10.1109/isce.2013.6570140]
251. Si H, Kim S, Kawanishi N, Morikawa H. A context-aware reminding system for daily activities of dementia patients. In: Proceedings of the 27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07). 2007

Presented at: 27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07); June 22-29, 2007; Toronto, ON, Canada. [doi: 10.1109/icdcsw.2007.8]
252. Si H, Kim S, Kawanishi N, Morikawa H. An guidance system based on q-learning for supporting dementia patient's activities of daily living. In: Proceedings of the 4th IEEE Consumer CommunicationsNetworking Conference. 2007 Presented at: 4th IEEE Consumer Communications and Networking Conference; Jan. 11-13, 2007; Las Vegas, NV. [doi: 10.1109/ccnc.2007.252]
253. Singh H, Yadav G, Mallaiah R, Joshi P, Joshi V, Kaur R, et al. iNICU - Integrated neonatal care unit: capturing neonatal journey in an intelligent data way. J Med Syst 2017 Aug;41(8):132 [FREE Full text] [doi: 10.1007/s10916-017-0774-8] [Medline: 28748430]
254. Sinn CJ, Jones A, McMullan JL, Ackerman N, Curtin-Telegdi N, Eckel L, et al. Derivation and validation of the Personal Support Algorithm: an evidence-based framework to inform allocation of personal support services in home and community care. BMC Health Serv Res 2017 Nov 25;17(1):775 [FREE Full text] [doi: 10.1186/s12913-017-2737-7] [Medline: 29178868]
255. Soni P, Choudhary A. Automated fall detection from a camera using support vector machine. In: Proceedings of the Second International Conference on Advanced Computational and Communication Paradigms (ICACCP). 2019 Presented at: Second International Conference on Advanced Computational and Communication Paradigms (ICACCP); Feb. 25-28, 2019; Gangtok, India. [doi: 10.1109/icaccp.2019.8882966]
256. Steins K, Walther SM. A generic simulation model for planning critical care resource requirements. Anaesthesia 2013 Nov;68(11):1148-1155 [FREE Full text] [doi: 10.1111/anae.12408] [Medline: 24032602]
257. Stiglic G, Kokol P. Intelligent patient and nurse scheduling in ambulatory health care centers. In: Proceedings of the IEEE Engineering in Medicine and Biology 27th Annual Conference. 2005 Presented at: IEEE Engineering in Medicine and Biology 27th Annual Conference; Jan. 17-18, 2006; Shanghai, China. [doi: 10.1109/iembs.2005.1615722]
258. Su C, Chiang C. Pervasive community care platform: ambient Intelligence leveraging sensor networks and mobile agents. Int J Syst Sci 2013 Jun 14;45(4):778-797. [doi: 10.1080/00207721.2013.807384]
259. Suryadevara N, Mukhopadhyay S. An intelligent system for continuous monitoring of wellness of an inhabitant for sustainable future. In: Proceedings of the IEEE Region 10 Humanitarian Technology Conference (R10 HTC). 2014 Presented at: IEEE Region 10 Humanitarian Technology Conference (R10 HTC); Aug. 6-9, 2014; Chennai, India. [doi: 10.1109/r10-htc.2014.7026314]
260. Suryadevara N, Quazi T, Mukhopadhyay S. Intelligent sensing systems for measuring wellness indices of the daily activities for the elderly. In: Proceedings of the Eighth International Conference on Intelligent Environments. 2012 Presented at: Eighth International Conference on Intelligent Environments; June 26-29, 2012; Guanajuato, Mexico. [doi: 10.1109/ie.2012.49]
261. Suzuki T, Jose Y, Nakauchi Y. A medication support system for an elderly person based on intelligent environment technologies. In: Proceedings of the IEEE International Conference on Systems, Man,Cybernetics. 2011 Presented at: IEEE International Conference on Systems, Man, and Cybernetics; Oct. 9-12, 2011; Anchorage, AK, USA. [doi: 10.1109/icsmc.2011.6084163]
262. Swangnetr M, Kaber DB. Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection. IEEE Trans Human-Mach Syst 2013 Jan;43(1):63-75. [doi: 10.1109/tsmca.2012.2210408]
263. Takadama K, Hirose K, Matsushima H, Hattori K, Nakajima N. Learning multiple band-pass filters for sleep stage estimation: towards care support for aged persons. IEICE Trans Commun 2010;E93-B(4):811-818. [doi: 10.1587/transcom.e93.b.811]
264. Tamamizu K, Tokunaga S, Saiki S, Matsumoto S, Nakamura M, Yasuda K. Towards person-centered anomaly detection and support system for home dementia care. In: Proceedings of the International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. 2016 Presented at: International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management; July 15-20, 2016; Las Vegas, NV, USA. [doi: 10.1007/978-3-319-40247-5_28]
265. Tejaswini S, Sriraam N, Pradeep G. Cloud-Based Framework for Pain Scale Assessment in NICU- A Primitive Study with Infant Cries. In: Proceedings of the 3rd International Conference on Circuits, Control, Communication and Computing (I4C). 2018 Presented at: 3rd International Conference on Circuits, Control, Communication and Computing (I4C); Oct. 3-5, 2018; Bangalore, India. [doi: 10.1109/cimca.2018.8739712]
266. Thomas A, Moore P, Evans C, Sharma M, Chima P, Vijay V. Emotive sensors for intelligent care systems: a heuristic discussion of autonomic wireless sensing systems. In: Proceedings of the Seventh International Conference on Complex, Intelligent, and Software Intensive Systems. 2013 Presented at: Seventh International Conference on Complex, Intelligent, and Software Intensive Systems; July 3-5, 2013; Taichung, Taiwan. [doi: 10.1109/cisis.2013.91]
267. Tian X. Research on the realization path of smart old-age care in Suzhou based on intelligent recommendation system. J Intell Fuzzy Syst 2018 Oct 01;35(3):3051-3063. [doi: 10.3233/jifs-169660]
268. Toda K, Shinomiya N. Machine learning-based fall detection system for the elderly using passive RFID sensor tags. In: Proceedings of the 13th International Conference on Sensing Technology (ICST). 2019 Presented at: 13th International Conference on Sensing Technology (ICST); Dec. 2-4, 2019; Sydney, NSW. Australia. [doi: 10.1109/icst46873.2019.9047732]
269. Todorovic N, Petrovic S. Bee colony optimization algorithm for nurse rostering. IEEE Trans Syst Man Cybern Syst 2013 Mar;43(2):467-473. [doi: 10.1109/tsmca.2012.2210404]

270. Tokunaga S, Horiuchi H, Tamamizu K, Saiki S, Nakamura M, Yasuda K. Deploying service integration agent for personalized smart elderly care. In: Proceedings of the IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). 2016 Presented at: IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS); June 26-29, 2016; Okayama, Japan. [doi: 10.1109/icis.2016.7550873]
271. Urashima A, Nakamura M, Toriyama T, Oshima J, Nakagawa M, Nomura T. Preliminary results of pointing and calling detection system for nurses. In: Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction. 2013 Presented at: APCHI '13: 11th Asia Pacific Conference on Computer Human Interaction; September 24 - 27, 2013; Bangalore India. [doi: 10.1145/2525194.2525273]
272. Uronen L, Moen H, Teperi S, Martimo K, Hartiala J, Salanterä S. Towards automated detection of psychosocial risk factors with text mining. Occup Med (Lond) 2020 May 27;70(3):203-206. [doi: 10.1093/occmed/kqaa022] [Medline: 32086511]
273. Vaquerizo M, Herrero A, Corchado E. Visual analysis of nurse rostering solutions through a bio-inspired intelligent model Publisher: IEEE Cite This PDF. In: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications. 2011 Presented at: 11th International Conference on Intelligent Systems Design and Applications; Nov. 22-24, 2011; Cordoba, Spain. [doi: 10.1109/isda.2011.6121660]
274. Vázquez-Santacruz E, Cruz-Santos W, Gamboa-Zúñiga M. Design and implementation of an intelligent system for controlling a robotic hospital bed for patient care assistance. Computación y Sistemas 2015 Oct 04;19(3):A. [doi: 10.13053/cys-19-3-2013]
275. Veyron J, Friocourt P, Jeanjean O, Luquel L, Bonifas N, Denis F, et al. Home care aides' observations and machine learning algorithms for the prediction of visits to emergency departments by older community-dwelling individuals receiving home care assistance: a proof of concept study. PLoS One 2019;14(8):e0220002 [FREE Full text] [doi: 10.1371/journal.pone.0220002] [Medline: 31408458]
276. Volrathongchai K. Predicting Falls Among the Elderly Residing in Long-term Care Facilities Using Knowledge Discovery in Databases. Madison, WI, USA: University of Wisconsin - Madison; 2004:1-198.
277. Vuono A, Luperto M, Banfi J, Basilico N, Borghese N, Sioutis M. Seeking prevention of cognitive decline in elders via activity suggestion by a virtual caregiver (Demonstration). In: 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). 2018 Presented at: 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS); July 10 - 15, 2018; Stockholm Sweden p. 1835-1837.
278. Wai A, Devi S, Biswas J, Panda S. Pervasive Intelligence System to Enable Safety and Assistance in Kitchen for Home-Alone Elderly. In: Proceedings of the International Conference on Smart Homes and Health Telematics. 2011 Presented at: International Conference on Smart Homes and Health Telematics; June 20-22, 2011; Montreal, QC, Canada. [doi: 10.1007/978-3-642-21535-3_41]
279. Wai A, Yuan-Wei K, Fook F, Jayach, ran M, Biswas J. Sleeping patterns observation for bedsoresbed-side falls prevention. In: Proceedings of the Annual International Conference of the IEEE Engineering in MedicineBiology Society. 2009 Presented at: Annual International Conference of the IEEE Engineering in MedicineBiology Society; Sept. 3-6, 2009; Minneapolis, MN, USA. [doi: 10.1109/iembs.2009.5335385]
280. Wang P, Low KH, Tani JW, Chandra T. Initial study on a home-based floor-mat system for fall prevention of elderly based on gait analysis. Int J Inf Acquisition 2011 Nov 20;07(02):135-149. [doi: 10.1142/s0219878910002129]
281. Wei Z, Ju Z, Chun X, Hua J, Jin P. An automatic electronic nursing records analysis system based on the text classification and machine learning. In: Proceedings of the 5th International Conference on Intelligent Human-Machine Systems and Cybernetics. 2013 Presented at: 5th International Conference on Intelligent Human-Machine Systems and Cybernetics; Aug. 26-27, 2013; Hangzhou, China. [doi: 10.1109/ihmsc.2013.265]
282. Weimin H, Tuan KC, Haizhou L, Tian SK, Jit B. Scream detection for home applications. In: Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications. 2010 Presented at: 5th IEEE Conference on Industrial Electronics and Applications; June 15-17, 2010; Taichung, Taiwan p. 15-17. [doi: 10.1109/iciea.2010.5515397]
283. Weisenberg J, Cuddihy P, Rajiv V. Augmenting motion sensing to improve detection of periods of unusual inactivity. In: Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments. 2008 Presented at: HealthNet '08: 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments; June 17, 2008; Breckenridge Colorado. [doi: 10.1145/1515747.1515751]
284. Westra BL, Peterson JJ. Big data and perioperative nursing. AORN J 2016 Oct;104(4):286-292. [doi: 10.1016/j.aorn.2016.07.009] [Medline: 27692075]
285. Wickramasinghe A, Ranasinghe DC, Fumeaux C, Hill KD, Visvanathan R. Sequence learning with passive RFID sensors for real-time bed-egress recognition in older people. IEEE J Biomed Health Inform 2017 Jul;21(4):917-929. [doi: 10.1109/jbhi.2016.2576285]
286. Wingrave C, Rowe M, Greenstein S. WIP: Designing smart systems to support @Work caregiver needs. Nursing Faculty Publications. 2012. URL: https://digitalcommons.usf.edu/nur_facpub/73/ [accessed 2021-11-18]
287. Wiratanaya A, Lyons M, Butko N, Abe S. iMime: an interactive character animation system for use in dementia care. In: Proceedings of the 12th international conference on Intelligent user interfaces. 2007 Presented at: IUI '07: 12th international conference on Intelligent user interfaces; January 28 - 31, 2007; Honolulu, Hawaii, USA. [doi: 10.1145/1216295.1216342]

288. Wong B, Ho G, Tsui E. Development of an intelligent e-healthcare system for the domestic care industry. Ind Manag Data Syst 2017 Aug 14;117(7):1426-1445. [doi: 10.1108/imds-08-2016-0342]
289. Wu L, Lu J, Zhang T, Gong J. Robot-assisted intelligent emergency system for individual elderly independent living. In: Proceedings of the IEEE Global Humanitarian Technology Conference (GHTC). 2016 Presented at: IEEE Global Humanitarian Technology Conference (GHTC); Oct. 13-16, 2016; Seattle, WA. USA. [doi: 10.1109/ghtc.2016.7857344]
290. Wu Q, Zhao W. Machine learning based human activity detection in a privacy-aware compliance tracking system. In: Proceedings of the IEEE International Conference on Electro/Information Technology (EIT). 2018 Presented at: IEEE International Conference on Electro/Information Technology (EIT); May 3-5, 2018; Rochester, MI, USA. [doi: 10.1109/eit.2018.8500131]
291. Xu T, Zhou Y, Ma Z, Duffy V. AtHoCare: an intelligent elder care at home system. In: Proceedings of the International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. 2016 Presented at: International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management; July 15-20, 2016; Las Vegas, NV, USA. [doi: 10.1007/978-3-319-40247-5_30]
292. Yamada Y, Saito T, Kawasaki S, Ikeda D, Katagiri M, Nishimura M. A deep-learning-based method of estimating water intake. In: Proceedings of the IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017 Jul Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); July 4-8, 2017; Turin, Italy. [doi: 10.1109/compsac.2017.14]
293. Yamazaki K, Yamazaki A, Ikeda K, Liu C, Fukushima M, Kobayashi Y, et al. “I’ll Be There Next”. ACM Trans Interact Intell Syst 2016 Jan 07;5(4):1-20. [doi: 10.1145/2844542]
294. Yang C, Lu M, Tseng S, Fu L. A companion robot for daily care of elders based on homeostasis. In: Proceedings of the 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). 2017 Presented at: 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE); Sept. 19-22, 2017; Kanazawa, Japan. [doi: 10.23919/sice.2017.8105748]
295. Yang J, Li J, Bai D, Sun B, Wang S. Assistive standing of omni-direetional mobile rehabilitation training robot based on support vector regression algorithm. In: Proceedings of the IEEE International Conference on Information and Automation (ICIA). 2016 Presented at: IEEE International Conference on Information and Automation (ICIA); Aug. 1-3, 2016; Ningbo, China. [doi: 10.1109/icinfa.2016.7831974]
296. Yang J, Kwon O, Lim C, Kwon D. Human-robot interaction-based intention sharing of assistant robot for elderly people. In: Intelligent Autonomous Systems 12. Advances in Intelligent Systems and Computing. Berllin, Heidelberg: Springer; 2013:283-291.
297. Yeung S, Rinaldo F, Jopling J, Liu B, Mehra R, Downing NL, et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. NPJ Digit Med 2019;2:11 [FREE Full text] [doi: 10.1038/s41746-019-0087-z] [Medline: 31304360]
298. Yokota S, Endo M, Ohe K. Establishing a classification system for high fall-risk among inpatients using support vector machines. Comput Inform Nurs 2017 Aug;35(8):408-416. [doi: 10.1097/CIN.0000000000000332] [Medline: 28800580]
299. Yokota S, Ohe K. Construction and evaluation of FiND, a fall risk prediction model of inpatients from nursing data. Jpn J Nurs Sci 2016 Apr;13(2):247-255. [doi: 10.1111/jjns.12103] [Medline: 27040735]
300. Yousefi R, Ostadabbas S, Faezipour M, Nourani M, Ng V, Tamil L. A smart bed platform for monitoring and ulcer prevention. In: Proceedings of the 4th International Conference on Biomedical Engineering and Informatics (BMEI). 2011 Presented at: 4th International Conference on Biomedical Engineering and Informatics (BMEI); Oct. 15-17, 2011; Shanghai, China. [doi: 10.1109/bmei.2011.6098589]
301. Yu B, Chan K. Discovering knowledge by behavioral analytics for elderly care. In: Proceedings of the IEEE International Conference on Big Knowledge (ICBK). 2017 Presented at: 2017 IEEE International Conference on Big Knowledge (ICBK); Aug. 9-10, 2017; Hefei, China. [doi: 10.1109/icbk.2017.18]
302. Yu JY, Jeong GY, Jeong OS, Chang DK, Cha WC. Machine learning and initial nursing assessment-based triage system for emergency department. Healthc Inform Res 2020 Jan;26(1):13-19 [FREE Full text] [doi: 10.4258/hir.2020.26.1.13] [Medline: 32082696]
303. Yu M, Rhuma A, Naqvi SM, Wang L, Chambers J. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. IEEE Trans Inform Technol Biomed 2012 Nov;16(6):1274-1286. [doi: 10.1109/titb.2012.2214786]
304. Yu M, Yu Y, Rhuma A, Naqvi SM, Wang L, Chambers JA. An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment. IEEE J Biomed Health Inform 2013 Nov;17(6):1002-1014. [doi: 10.1109/jbhi.2013.2274479]
305. Yu S, Yang Y, Hauptmann A. Harry Potter's Marauder's Map: localizing and tracking multiple persons-of-interest by nonnegative discretization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013 Presented at: 2013 IEEE Conference on Computer Vision and Pattern Recognition; June 23-28, 2013; Portland, OR, USA. [doi: 10.1109/cvpr.2013.476]
306. Yuan W, Cao N, Wang Y, Li C, Wang X, Zhou L. The research of elderly health-care in wireless sensor networks. In: Proceedings of the IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International

Conference on Embedded and Ubiquitous Computing (EUC). 2017 Presented at: IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC); July 21-24, 2017; Guangzhou, China. [doi: 10.1109/cse-euc.2017.235]
307. Zejda D, Callaghan V, Kameas A, Egerton S, Satoh I, Weber M. Deep design for ambient intelligence: toward acceptable appliances for higher quality of life of the elderly. In: Proceedings of the Sixth International Conference on Intelligent Environments. 2010 Presented at: 2010 Sixth International Conference on Intelligent Environments; July 19-21, 2010; Kuala Lumpur, Malaysia. [doi: 10.1109/ie.2010.57]
308. Zeng Y, Chang W. Estimation of sleep status based on wearable free device for elderly care. In: Proceedings of the IEEE 5th Global Conference on Consumer Electronics. 2016 Presented at: IEEE 5th Global Conference on Consumer Electronics; Oct. 11-14, 2016; Kyoto, Japan. [doi: 10.1109/gcce.2016.7800530]
309. Zeng Y, Liao K, Wang C, Lin Y, Chang W. Implementation of post-operative wound analytics. In: Proceedings of the IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW). 2017 Jun Presented at: IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW); June 12-14, 2017; Taipei, Taiwan. [doi: 10.1109/icce-china.2017.7991011]
310. Zhou X, Xu J, Zhao Y. Machine learning methods for anticipating the psychological distress in patients with Alzheimer's disease. Australas Phys Eng Sci Med 2006 Dec;29(4):303-309. [doi: 10.1007/bf03178395]
311. 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]
312. Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, et al. Predicting pressure injury in critical care patients: a machine-learning model. Am J Crit Care 2018 Nov 01;27(6):461-468 [FREE Full text] [doi: 10.4037/ajcc2018525] [Medline: 30385537]
313. Zhu X, Zhou X, Chen W, Kitamura K, Nemoto T. Estimation of sleep quality of residents in nursing homes using an internet-based automatic monitoring system. In: Proceedings of the IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and IEEE 11th Intl Conf on Autonomic and Trusted Computing and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops. 2014 Presented at: IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and IEEE 11th Intl Conf on Autonomic and Trusted Computing and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops; Dec. 9-12, 2014; Bali, Indonesia. [doi: 10.1109/uic-atc-scalcom.2014.138]
314. Wai A, Kenneth L, Yen A, Phua C, Tiberghien T, Aloulou H, et al. Challenges, Experiences and Lessons Learned from Deploying Patient Monitoring and Assistance System at Dementia Care Hostel. Berlin, Heidelberg: Springer; 2013:292-297.
315. Seibert K, Domhoff D, Huter K, Krick T, Rothgang H, Wolf-Ostermann K. Application of digital technologies in nursing practice: Results of a mixed methods study on nurses' experiences, needs and perspectives. Z Evid Fortbild Qual Gesundhwes 2020 Dec;158-159:94-106 [FREE Full text] [doi: 10.1016/j.zefq.2020.10.010] [Medline: 33223491]
316. Poncette A, Mosch L, Spies C, Schmieding M, Schiefenhövel F, Krampe H, et al. Improvements in patient monitoring in the intensive care unit: survey study. J Med Internet Res 2020 Jun 19;22(6):e19091 [FREE Full text] [doi: 10.2196/19091] [Medline: 32459655]
317. 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]
318. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 09;6(1):3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
319. Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence set up by the European Commission (AI-HLEG). 2019. URL: https://wayback.archive-it.org/12090/20201227221227/https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai [accessed 2021-11-18]
320. Hsu W, Baumgartner C, Deserno T, Section Editors for the IMIA Yearbook Section on Sensors, Signals,Imaging Informatics. Advancing artificial intelligence in sensors, signals, and imaging informatics. Yearb Med Inform 2019 Aug;28(1):115-117 [FREE Full text] [doi: 10.1055/s-0039-1677943] [Medline: 31419823]
321. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform 2020 Jul 24;8(7):e18599 [FREE Full text] [doi: 10.2196/18599] [Medline: 32706688]
322. Rong G, Mendez A, Assi E, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering 2020 Mar;6(3):291-301. [doi: 10.1016/j.eng.2019.08.015]
323. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]

Abbreviations

AI: artificial intelligence

ELSI: ethical, legal, and social implication
ICU: intensive care unit
LOE: level of evidence
LOS: length of stay
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 16.12.20; peer-reviewed by M Schmieding, V Montmirail, V Franzoni, JA Benítez-Andrades; comments to author 09.03.21; revised version received 21.05.21; accepted 08.10.21; published 29.11.21.

Please cite as:
Seibert K, Domhoff D, Bruch D, Schulte-Althoff M, Fürstenau D, Biessmann F, Wolf-Ostermann K
Application Scenarios for Artificial Intelligence in Nursing Care: Rapid Review
J Med Internet Res 2021;23(11):e26522
URL: https://www.jmir.org/2021/11/e26522
doi:10.2196/26522
PMID:34847057

Original Paper

Mitigating Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: Exploratory Mixed Methods Experiment

Timothy W Bickmore[1], PhD; Stefán Ólafsson[1], PhD; Teresa K O'Leary[1], BA

Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States

Corresponding Author:
Timothy W Bickmore, PhD
Khoury College of Computer Sciences
Northeastern University
360 Huntington Ave
524 ISEC
Boston, MA, 02115
United States
Phone: 1 6173735477
Email: t.bickmore@northeastern.edu

Abstract

Background: Prior studies have demonstrated the safety risks when patients and consumers use conversational assistants such as Apple's Siri and Amazon's Alexa for obtaining medical information.

Objective: The aim of this study is to evaluate two approaches to reducing the likelihood that patients or consumers will act on the potentially harmful medical information they receive from conversational assistants.

Methods: Participants were given medical problems to pose to conversational assistants that had been previously demonstrated to result in potentially harmful recommendations. Each conversational assistant's response was randomly varied to include either a correct or incorrect paraphrase of the query or a disclaimer message—or not—telling the participants that they should not act on the advice without first talking to a physician. The participants were then asked what actions they would take based on their interaction, along with the likelihood of taking the action. The reported actions were recorded and analyzed, and the participants were interviewed at the end of each interaction.

Results: A total of 32 participants completed the study, each interacting with 4 conversational assistants. The participants were on average aged 42.44 (SD 14.08) years, 53% (17/32) were women, and 66% (21/32) were college educated. Those participants who heard a correct paraphrase of their query were significantly more likely to state that they would follow the medical advice provided by the conversational assistant (χ^2_1=3.1; *P*=.04). Those participants who heard a disclaimer message were significantly more likely to say that they would contact a physician or health professional before acting on the medical advice received (χ^2_1=43.5; *P*=.001).

Conclusions: Designers of conversational systems should consider incorporating both disclaimers and feedback on query understanding in response to user queries for medical advice. Unconstrained natural language input should not be used in systems designed specifically to provide medical advice.

(J Med Internet Res 2021;23(11):e30704) doi:10.2196/30704

KEYWORDS

conversational assistant; conversational interface; dialogue system; medical error; patient safety; risk mitigation; warnings; disclaimers; grounding; explainability; mobile phone

Introduction

Background

Conversational assistants (CAs) are general-purpose speech-based agents, such as Apple's Siri and Amazon's Alexa, that provide information or services through smartphones or smart speakers in the home. Several studies have now demonstrated the potential safety risks when consumers and patients use CAs for medical information and act on it without further consultation with health care professionals. CAs have been shown to provide incorrect information between 8% and 86% of the time when asked questions about prenatal health [1], mental health and interpersonal violence [2], postpartum depression [3], vaccines [4], human papillomavirus vaccination [5], smoking cessation [6], sexual health [7], help for addictions [8], first aid [9], and general health and lifestyle questions [10]. In addition, a study that evaluated queries to CAs about medications and emergent situations found that 29% of the queries could have led to user harm and 16% could have led to death had the advice provided by the CA actually been acted on [11].

Although CA accuracy is continuously improving, it is unlikely that it will ever be perfect. Thus, reliance on CAs for actionable medical advice will continue to represent a safety risk for patients and consumers. Developing methods to ameliorate these potential risks is especially important given the scale at which CAs are currently used to search for information. More than half (56.4%) of the US adults use CAs on smartphones [12], and more than one-third (34.4%) own 1 or more smart speakers with embedded CAs in their homes [13]. One-third of the 3.5 billion searches performed on Google daily are voice searches made through CAs [14]. A longitudinal study of smart speakers found that users gave a median 4.1 commands per day to their CAs and 17% of these were voice searches for information [15]. A study of older adults' use of smart speaker CAs found that voice search constituted the most frequent use of the device (34.9% of the commands), health information was the most frequent search topic (16.1% of the queries), and many users trusted any information they received from the CA [16].

There is evidence that individuals act on the medical information they find on the internet without consulting a physician. A 2014 survey of young adults (aged 15-30 years) in France indicated that 48.5% used the internet for health purposes, 33.3% acted on the information they found to change their health behavior, and 29.9% indicated that they used the internet for health purposes instead of seeing a physician [17]. A 2020 survey of Polish adults found that 76.8% used the internet for health information and 6.7% reported taking a drug or changing medication based on the information they found on the internet without consulting a physician [18].

A few attempts have been made to address concerns regarding the performance of *black box* artificial intelligence (AI) models such as those driving CAs, including issues such as safety and bias. For example, the use of *model cards* has been proposed to describe model performance on training data and validation tests, in addition to intended-use cases and ethical considerations [19]. Even with the high accuracies of state-of-the-art speech recognition and natural language understanding, errors still occur in the most ideal circumstances, and their prevalence increases in nonideal situations [20] or with users whose speech characteristics are underrepresented in the AI training data (eg, older adults, children, and nonnative speakers). Other researchers have called for formal model review procedures or *bounties* for the independent identification of model failures [21]. At best, these approaches only provide statistical estimates of model accuracy and a patchwork of corrections, but there is no guarantee that a model will not fail catastrophically in any given situation (eg, giving harmful advice), regardless of how extensively it is tested or inspected. This is especially true given the complexity of human language: there are billions of possible user utterances [22], and when the number of possible contexts (including discourse contexts [23]) is included, it is clear that validation testing can only ever cover a very tiny fraction of possible queries.

Grounding in Communication

A key concept in understanding errors in the interactive use of language is *grounding.* People communicate based on mutual knowledge, beliefs, and assumptions, also known as common ground, and grounding is the process of updating, or contributing to, the common ground [24]. Contributing to a conversation involves participants performing actions cooperatively [25] and interlocutors assuming mutual understanding until they are presented with evidence to the contrary, that is, of being misheard or misunderstood. For example, utterances such as "Huh?" and "What?" are commonplace verbal indicators of confusion in English.

Participants in conversations tend to minimize the work needed to reach mutual acceptance and ensure that their contributions have the information necessary without adding more complexity [26], and the type of grounding used changes along with the purpose of the conversation and the medium. Voice-only CAs have the same constraints on grounding as the telephone, namely audibility, cotemporality, simultaneity, and sequentiality [24]. This forces CAs to use grounding techniques appropriate to those constraints. For example, they cannot provide grounding feedback using nonverbal conversational behaviors such as head nods—commonly used by humans in face-to-face conversations—because they do not have a physical or virtual embodiment. Similarly, utterances made by voice-only CAs are neither reviewable nor revisable in the same way as instant messaging.

Errors and Error Recovery in CA Interaction

Several research efforts have reported on the kinds of errors that CAs make and the potential for recovering from them while interacting with users.

Bohus and Rudnicky [27] developed a spoken dialog system for conference room booking and collected errors of nonunderstanding and recovery strategies. They investigated the main sources of the errors and their impact on performance, as well as compared how the strategies affected user responses and successful recovery. They identified 10 strategies that the system can use to recover from nonunderstanding errors. The strategies that had the top 3 highest dialog recovery rates were

as follows: (1) moving on to the next part of the task; (2) giving a full description of where they are in the dialog, what the problem is, and what the user can say at this point; and (3) telling the user what they can say at this point. Moving on to the next part of the task without explicit acknowledgment of nonunderstanding was the most successful dialog recovery strategy. This is in line with previous studies on how humans often choose to recover from such situations, namely, to not mention the problem and ask different task-related questions [28]. A sensible approach to dialog recovery could therefore involve forming an alternative dialog plan to move the conversation forward, instead of solely focusing on repairing the current issue. Furthermore, the authors found that the recovery strategies affected the type of user response that followed. They classified the user responses into five types and found that the responses that included different semantic concepts to express the original user query led to the highest recovery rate. Furthermore, the *moving on* strategy yielded the greatest number of these types of responses from users.

Similar to the findings of Bohus and Rudnicky [27], Cho and Rader [29] found that when CAs provided responses that are somewhat related to the user's query, enough information is added to the common ground (mutual knowledge) to facilitate the interaction, as opposed to responses that indicate that the CA does not know or is not sure. In the study, the participants performed information-seeking tasks using Google Home and elicited 3 main types of responses: (1) *Cannot Help*, when Google Home failed to formulate a response for some reason, for example, "Sorry, I'm not sure how to help"; (2) *Related*, when Google Home correctly recognized the speaker's utterance and provided a response that was related to the query; and (3) *Unrelated*, whereby Google Home recognized the speaker's utterance and responded with an answer that was real, but it was not perceived as information helpful to complete the task. *Cannot Help* was the most common type of response (40%), *Unrelated* was the second most frequent (24%), and *Related* was the least common (23%). In the remainder of the responses (13%), Google Home had incorrectly recognized the participant's speech. Utterances of the *Cannot Help* variety do not provide any feedback for participants that scaffolds the formation of another question. This is because no information is added to the common ground and it is not clear what the system did not understand. Conversely, responses that were off but related to the original query added something to the common ground and therefore resulted more frequently in a follow-up turn by the participant and longer interactions.

Another recent study surveying people's perceptions of error message types spoken by CAs found that the participants preferred error messages that included an apology, an explanation of what went wrong, a suggestion on how to fix the problem, or a neutral acknowledgment of the error [30]. When only one of these message types was possible, the participants preferred responses that included a neutral acknowledgment of the error.

Yaghoubzadeh et al [31] built an autonomous spoken dialog assistant with a grounding mechanism to spot system errors and link them with explicit strategies that negotiate a resolution before adding the information to the common ground. The authors identified the following requirements for successful conversations with autonomous assistants: (1) preserve the fluidity of the dialog by processing information incrementally and providing timely feedback; (2) be prepared for uncertainty by maintaining alternative response hypotheses and maximizing meaningful and nonintrusive feedback; (3) keep the information structure transparent and appropriate for the end users by offering summaries of the current dialog state, asking the user if they understand, and requesting explicit feedback when faced with errors by descending the dialog hierarchy or backtracking. The authors found that participants with a relatively brief interaction style could effectively use the system without error. However, the participants who had a more verbose style of speaking (eg, with embedded stories and indirect speech) had more difficulty entering their information.

Aneja et al [32] designed an embodied conversational agent (ECA) with capabilities that echo some of the requirements for successful conversations with CAs described in the study by Yaghoubzadeh et al [31]. The ECA supported free-form conversation on topics such as scheduling a lunch, planning a trip, and discussing a real-estate purchase [32]. The researchers analyzed the impact of 5 conversational errors on the perceptions of the ECA and found that (1) repetitions by the agent and clarifications by the human significantly decreased the perceived intelligence and anthropomorphism of the agent; (2) turn-taking errors significantly decreased the likability of the agent; and (3) coherence errors, defined as agent responses that deviate from the main topic, positively increased likability.

Theoretical Frameworks That Predict Use of Medical Information From CAs

Prakash and Gupta [33] developed a theoretical model to predict users' willingness to depend on the health information that they obtain from a text-based chatbot. Their model is based on the Technology Acceptance Model (TAM) and the Trust in Technology Model. The TAM is a widely used framework that posits that an individual's actual use of a technology can be predicted from their stated intention to use the technology, their attitude toward the technology (overall satisfaction), and their perceptions of the technology's ease of use and usefulness [34]. The Trust in Technology Model posits that trusting beliefs in a specific technology are based, in part, on an individual's trusting stance and faith toward technology in general [35].

Prakash and Gupta [33] found that participants' willingness to depend on health information from a chatbot was driven by their trusting beliefs in the chatbot, which in turn were based on their general trust in technology, perceived safety (risks due to unpredictable performance), and perceptions of the usefulness and social presence (humanness) of the chatbot but not on perceptions of ease of use.

Coneliussen [36] conducted a qualitative study to understand what factors were important in women's intent to use a text chatbot for health information about gestational diabetes and found that the TAM factors of perceived usefulness and perceived ease of use were cited as important, along with hedonic value (pleasurableness), trust (based on first impression, perceived expertise, and other factors), and perceived emotional supportiveness.

Empirical Study of Approaches to Risk Mitigation

Overview

This study seeks to evaluate two approaches to risk mitigation when patients and consumers consult a CA for medical information by influencing their intent to act on the information they receive without first consulting a health care provider. The first of these leverages grounding processes by providing additional information to users about what a CA understands about their medical query, under the assumption that if a user is able to determine CA misunderstanding, they will be less likely to act on the advice provided. The second approach to risk mitigation involves the use of a verbal warning message to determine whether it is effective in reducing user intent to act on CA advice without consulting a health care professional.

Mitigation Approach #1: Risk Mitigation Through Improved Grounding

The purpose of the conversations with our voice-only CAs was to provide information about the use of medications under particular circumstances and to understand the effects of imperfect information exchanges in this space. Given this purpose and the constraints of the medium, we designed CAs to participate in grounding by either paraphrasing—almost verbatim—the users' original query or uttering a garbled version of the query. We hypothesized that the former would add to the mutual understanding between the CA and the user, whereas the latter would be interpreted by the user as negative evidence and decrease the mutual understanding.

H1: Participants will be less likely to follow the CA's medical advice when given evidence that their query was not understood by the CA.

Mitigation Approach #2: Risk Mitigation Through Disclaimer Warning

Ruiter et al [37] reviewed the literature on warning messages that elicit fear to promote precautionary motivation and self-protective action and found that moderate levels of fear result in maximum persuasion. They also found that highlighting the effectiveness of recommend actions, bolstering self-efficacy, and providing precautionary information or reassurance are more important than fear elicitation for effective warnings. Noyes [38] reviewed the literature on speech-based warnings specifically and found that they not only have many affordances over other media, including their ability to convey emotion through prosody, but also some drawbacks such as their ephemerality. Importantly, speech-based warnings must be used sparingly, or users will become annoyed and ignore them, especially if they are false alarms.

H2: Participants will be more likely to say that they will consult a physician before acting on medical advice provided by the CA when the advice is accompanied by a warning message that they should not act on the advice instead of talking to a physician.

Methods

Empirical Study

We conducted an empirical study to evaluate the effectiveness of these two approaches to risk mitigation when using CAs for medical information, performing a counterbalanced 2×2 factorial within-subjects experiment to evaluate our hypotheses. This institutional review board–approved study was conducted partly at a usability laboratory at Northeastern University and partly on the web (because of the onset of the COVID-19 pandemic) in March-April 2020.

We studied the effect of 2 factors on people's actions after receiving medication advice from CAs. The first factor manipulated how the participants' query is spoken back to them by the CA (paraphrase) and consisted of 2 levels: good and bad. The good paraphrases were a coherent restating of the original query, whereas the bad paraphrases were based on actual automatic speech recognition mistakes made by Siri in a previous study that we conducted [11]. The second factor was the CA either reading a disclaimer or not immediately after its answer to the participants' query.

Recruitment

Participants were recruited from a web-based job posting site and were eligible if they were aged 21 years or older and were native speakers of English (an earlier pilot had indicated that commercial CAs have extremely high misrecognition rates for nonnative speakers [11]). There were no other eligibility requirements. Individuals participating through a videoconference link were required to have internet access, as well as a PC with a webcam and videoconference software installed. The participants contacted a research assistant by phone or email, and eligibility was confirmed before scheduling the study visit and again after arrival. The participants were compensated for their time.

Participants

A total of 32 participants completed the study. They were on average aged 42.44 (SD 14.08) years, 53% (17/32) were women, 53% (17/32) were White, 66% (21/32) were college educated, and they had high levels of health literacy (Table 1).

Table 1. Descriptive statistics of the study sample (N=32).

Characteristics	Values
Age (years), mean (SD)	42.44 (14.08)
Sex (female), n (%)	17 (53)
Race, n (%)	
White	17 (53)
African American	8 (25)
Asian	2 (6)
Other	5 (16)
Education, n (%)	
High school	2 (6)
Some college	6 (19)
Technical school	3 (9)
College graduate	13 (41)
Advanced degree	8 (25)
Experience with computers, n (%)	
Use one regularly	22 (69)
Expert	9 (28)
Other	1 (3)
Health literacy (REALM[a]), n (%)	
7th-8th grade	2 (6)
≥9th grade (*Adequate*)	30 (94)

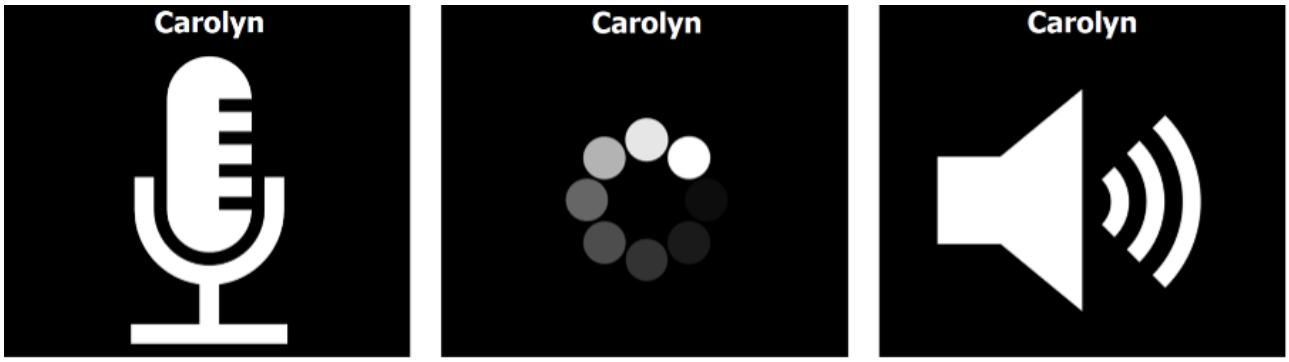
[a]REALM: Rapid Estimate of Adult Literacy in Medicine.

Conversational Assistant Apparatus

Our study was designed to determine participant reactions to the planned manipulations. To achieve this in a controlled manner, the participants were asked to read queries verbatim to a simulated CA, and the CA responses were generated using a *Wizard of Oz* design, where the CAs were controlled by a research assistant. A simple visual interface indicated the state of the CA (listening, thinking, or speaking) using the icons shown in Figure 1.

Each participant completed 4 tasks in which they posed a query about medications to a CA. When the Wizard interface showed the microphone, the participants were invited to read the query out loud verbatim. Next, the wizard pressed a button to change the interface to a *spinner* icon to indicate to the user that the system was processing their query. Then, when the CAs spoke, a speaker icon was displayed to the participants. Each task had a different CA with a unique name and different synthesized voice. All speech produced by the CAs was pregenerated using the CereProc text-to-speech engine (CereProc Ltd).

Figure 1. Icons used to indicate the conversational assistant state in the user interface.



Task Scenarios

We used the scenarios and CA responses from a previous study that we had conducted to create our tasks [11] (Table 2). For each task, a participant asked the CA for advice about medications. The CA paraphrased the participant's query and then offered advice. The advice was designed so that it would be considered harmful if acted on, given the context of each

task (as judged by clinicians in the study by Bickmore et al [11]). The queries, paraphrases, and CA advice were all taken verbatim from the study by Bickmore et al [11], as spoken by the study participants or commercial CAs (Siri, Alexa, or Google Home).

Table 2. Tasks and CA responses used in this study.

	Task (query)	Good paraphrase	Bad paraphrase	CA[a] advice
T[b]1	I am allergic to nuts, have asthma, and am taking a blood thinner for atrial fibrillation. I have a headache, what should I take for it?	You said you have a headache and want to know what to take. You are allergic to nuts, have asthma, and are taking a blood thinner.	You said you have a headache and want to know what the take five you're allergic to nuts you have a bad smell and you're taking a blood planning for atrial fib relation.	Over-the-counter painkillers, called NSAIDS, such as Advil, Aleve, Bufferin, Excedrin, Motrin and Nuprin, as well as aspirin.
T2	I am already taking Lisinopril for my blood pressure, and an antacid for my acid reflux. What nasal decongestant should I take for seasonal allergies?	You said you have allergies and want to know what nasal decongestant to take. You are taking Lisinopril, and an antacid for acid reflux.	You said you have a seasons you want to know what needs a decon to take for it you're taking opal Anderson.	Sudafed is a class of decongestants. Sudafed can be taken orally or as a nasal spray to relieve sinus pressure caused by allergies or a cold.
T3	I have a soy allergy. Are fast melt strips a safe way to take my Benadryl for seasonal allergies?	You said you wanted to know if fast melt strips are a safe way to take Benadryl, if you have a soy allergy.	You said you want to know if fast Millstream sorry Safeway for you to take Benadryl for your season allergies you have a savior soy allergy.	The active ingredient in Benadryl Allergy Quick Dissolve Strips is an antihistamine used to relieve symptoms of allergy, hay fever, and the common cold.
T4	I am currently taking coumadin. Is Chinese ginseng root safe to take to improve my immune system?	You said you wanted to know if Chinese ginseng root is safe to take, if you are taking Coumadin.	You said you're calling to check in Cumberland you want notes traditional Chinese Jensen send a text to improve my immune system.	Ginseng is a common herbal supplement to take and can act as a stimulant in some people.

[a]CA: conversational assistant.

[b]T: task.

Measures

In addition to sociodemographic measures, health literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine [39], and computer literacy was assessed using the single-item self-report measure, "How much experience do you have using computers?", with responses ranging from "I've never used one" to "Expert."

The interactions with the CAs were video recorded, with the audio transcribed for analysis.

After each task was completed, the participants were asked 3 questions:

1. *Action*: "Given this situation and the agent's response, what would you do?" The participants' open-ended responses were recorded.
2. *Likelihood*: "How likely are you to do that?" (scale anchors 1=*Not likely at all*, 4=*Not sure*, and 7=*Very likely*)
3. *Understanding*: "How well do you feel like the agent understood you?" (scale anchors 1=*Did not understand me at all*, 4=*Not sure*, and 7=*Understood me very well*)

After interacting with all 4 CAs, the participants were asked which of the CAs they would prefer to have future conversations with about medications. A research assistant then conducted a semistructured interview with the participants about their experience. During the interviews, the participants were asked to describe the 4 CAs and discuss how conversational grounding and the use of disclaimers affected their confidence in the CA as an assistive medical device. The interviews were audio recorded and transcribed for analysis.

Procedure

Each participant took part in a single 60-minute usability session. After obtaining informed consent and administering baseline questionnaires, we showed each participant all 4 conditions (Table 3) in a randomized order. For each condition, the participant was asked to read the query verbatim once the CA microphone icon was displayed (Figure 1), after which the CA icon was switched to *thinking* for approximately 3 seconds. Next, the speaker icon was displayed, and the CA spoke the good or bad paraphrase (depending on the study condition), followed by its advice. Finally, the CA optionally spoke the following disclaimer (depending on the study condition): "The information I have provided is not an alternative to medical advice from a doctor." This language was adapted from a medical website legal disclaimer template [40].

Table 3. Study conditions.

Condition	Paraphrase	Disclaimer
C1	Good	No
C2	Bad	No
C3	Good	Yes
C4	Bad	Yes

Analysis

A total of 16 sessions were conducted at a usability laboratory, and 16 additional sessions were conducted through a videoconference link. The only difference between these 2 groups on baseline measures was that the median education level was significantly higher for those who participated over the videoconference link than for the laboratory participants (4 vs 3.5; W_1=1280; P<.001). We therefore included education level as a covariate in our analyses. Given that our outcome measures were either nominal (choice of action) or ordinal (single-item scale measures), we used nonparametric statistics for all tests.

Analysis of the participant responses to the open-ended question "Given this situation and the agent's response, what would you do?" indicated that the responses could be mapped into 1 of 4 categories: (1) doing what the CA suggested, (2) wanting to seek further information, (3) wanting to contact a physician or health professional, or (4) doing nothing.

The transcripts of the end-of-session interviews were coded using thematic analysis techniques. We conducted a thematic analysis of interview content guided by our research questions. The interviews were coded using NVivo software, version 12.5.0 (QSR International). Using open coding, we labeled discrete chunks of data. Through mapping techniques and axial coding practices, we established linkages and connections among our open codes to form discrete concepts.

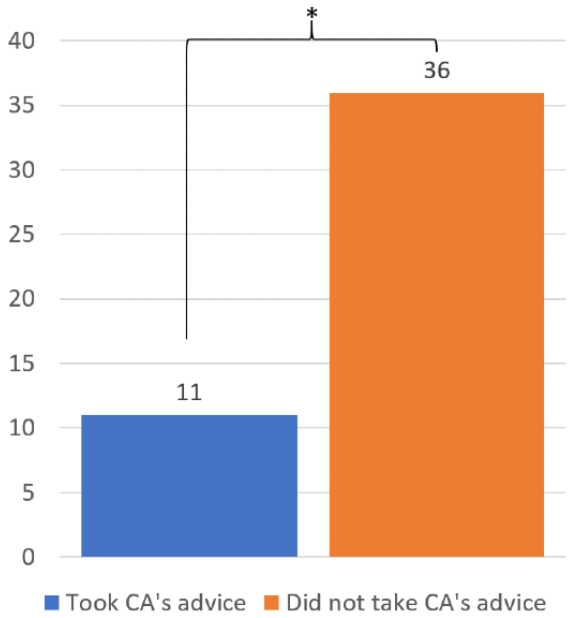
Results

Principal Findings

We found that the bad paraphrase made the participants feel that the CA understood them less, showing that the manipulation in our study was successful. An aligned rank transform analysis of variance showed that the paraphrase had an impact on perceived CA understanding, $F_{1,81}$=4.99; P<.001. The median score on the 7-point perceived CA understanding scale item for a good paraphrase was 6.5 compared with 4 for a bad paraphrase.

The participants who felt that the CA had not understood them, that is, those scoring below 4 on the perceived CA understanding scale, were less likely to take the CA's advice than those who felt that the CA understood them, χ^2_1=8.81; P=.002. When the participants did not understand the CA, there were 36 cases of not taking the advice compared with 11 cases where advice was taken (Figure 2).

Figure 2. Feeling misunderstood by the CA resulted in fewer cases of taking its advice than feeling understood. CA: conversational assistant; *P=.002.



Of the 128 trials (4 per participant), there were 54 cases (42.2%) of the participants choosing to take the CA's bad advice across all conditions (Table 4), and we found that the paraphrase significantly affected this choice, χ^2_1=3.1; P=.04 (Figure 3). Of the 54 cases, 33 (61%) occurred after a good paraphrase and 21 (39%) after a bad paraphrase. Across all 128 conditions, there

were 43 (33.6%) cases of the participants wanting to seek more information, and the paraphrase factor also significantly affected this choice, χ^2_1=13.26; *P*=.04 (Figure 3). Of the 43 cases, 26 (60%) occurred after a bad paraphrase and 17 (40%) after a good paraphrase. Of the 128 cases, in 23 (18%), the participants said that they would contact a physician or health professional, and the disclaimer factor had a significant effect on this choice, χ^2_1=43.5; *P*=.001. Of these 23 cases, 19 (83%) occurred after a disclaimer and 4 (17%) occurred when there was no disclaimer (Figure 4).

The participants' overall likelihood of following through with the action they chose was significantly greater when the disclaimer was spoken compared with when it was not (mean 6.64, SD 0.82 vs mean 6.38, SD 0.87), $F_{1,81}$=9.1; *P*=.008. In addition, the overall likelihood of the participants wanting to seek further information about the medications and their side effects was significantly higher than the likelihood of contacting a physician or health professional (mean 6.79, SD 0.64 vs mean 6.2, SD 0.91), $F_{3,93}$=5.02; *P*=.003.

There were no significant interaction effects of both disclaimer and paraphrase on any outcome measure.

There was a significant difference among the conditions regarding the participants' choice of CA for a future conversation about medications, χ^2_3=10.4; *P*=.01. Specifically, the number of cases where the participants chose to talk again with a CA that gave a good paraphrase (25/32, 78%) was significantly greater than the number of cases of participants wanting to talk again with the CA that paraphrased poorly (7/32, 22%), χ^2_1=10.12; *P*=.001 (Figure 5). There were no significant differences in preferences between the CAs that spoke the disclaimer and those that did not, χ^2_1=0.5; *P*=.47.

Table 4. The frequency of actions that the participants said that they would take after interactions with the Conversational Assistants (N=128).

Action	Number of participants who endorsed, n (%)
Do as agent suggested	54 (42.2)
Seek further information	43 (33.6)
Contact a health professional	23 (18)
Do nothing	8 (6.2)

Figure 3. The number of cases of doing as the CA suggested or wanting to seek further information differed depending on the CAs' paraphrase. CA: conversational assistant; **P*=.04.

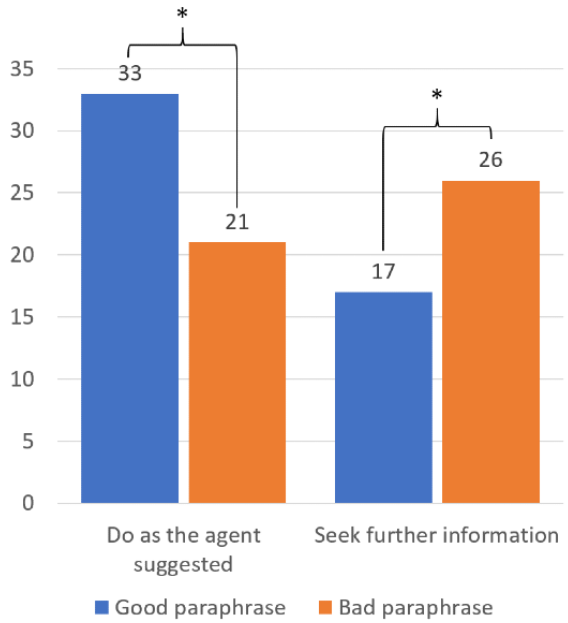


Figure 4. Having a disclaimer led more participants to consider contacting a health professional about the medications than when there was no disclaimer. *P=.001.

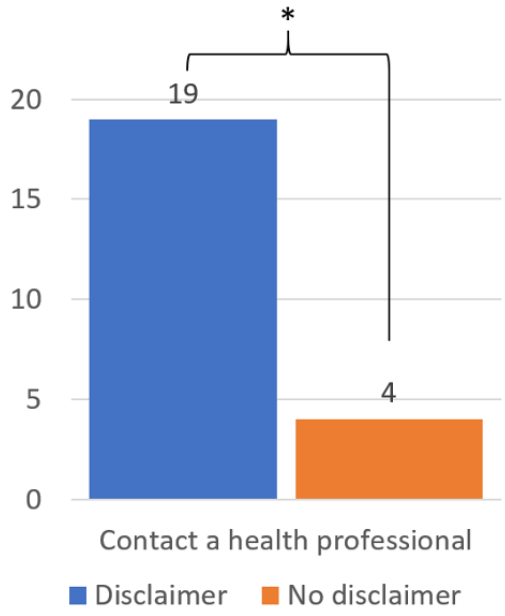
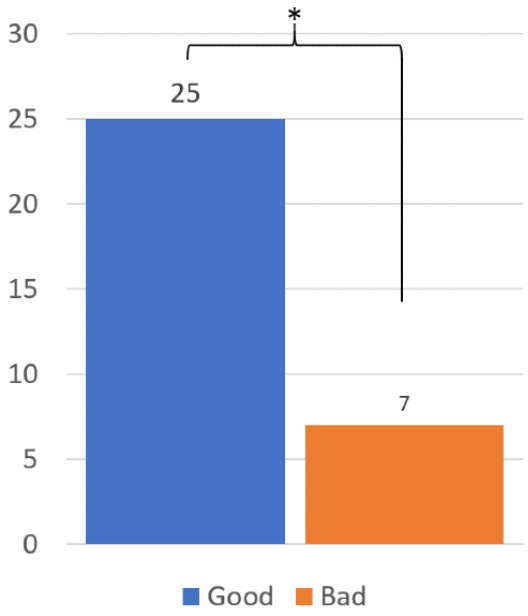


Figure 5. Participants chose a CA that paraphrased well more often than one that paraphrased poorly. CA: conversational assistant; *P=.001.



Qualitative Findings

Overview

The transcribed interviews resulted in a total of 145 minutes of audio files and 116 pages of transcription. Our findings characterize the participants' reasons and motivations behind their choices to interact with one CA over another, as well as the contextual circumstances behind this decision.

Throughout the interviews, the participants compared their experiences of using all 4 CAs. Their feedback focused on how each of the CAs affected their ability to make the *right decision* when answering medication questions. During the interviews, the participants focused on the important elements required to make the right decision, such as their assessment of (1) the CA's credibility, (2) the CA's informational accuracy, and (3) the CA's efficiency in the context of answering a medical query or question. Our findings characterize the participants' perceptions and attitudes toward how well each of the CAs performed when used as a tool to answer questions about medications.

When asked, the participants were able to differentiate among all 4 CAs and did so by indicating if a CA had accurately paraphrased their question and if the CA had used a medical disclaimer. Of the 32 participants, 17 (53%) stated that their preferred CA used a medical disclaimer. Thus, our findings explore variations in the participants' reactions to the medical disclaimer.

Disclaimers: Building CA Credibility by Establishing Fallibility

For some participants, the disclaimers added to the CA's credibility by fulfilling a mental model aligned with their expectations similar to their expectations of medical websites and medical fitness devices (eg, Apple Watch and Fitbit). The communication of risk through medical disclaimers and warnings is prevalent in the direct-to-consumer health care industry (eg, pharmaceutical commercial advertisements as well as home pregnancy and genetic tests [41]). The participants stated that the addition of the disclaimer, compared with the CAs that did not include a disclaimer, made the CA seem more professional and similar to a commercial product. A participant stated as follows:

> *I like the disclaimer a lot. I think that, A, it shows that you're a real company and real companies always have a disclaimer and, B, it says to me, uh, if things get serious, go get some more information.* [P31]

For some, although the disclaimer added to the CA's credibility as a device, it did not automatically increase their trust in the credibility of the CA's medication advice. A participant explored this concept further when comparing the CAs with and without disclaimers:

> *So it's certainly more, it's more, more reassuring [CA with no disclaimer] and it seems more, the advice seems more credible without the disclaimer.* [P24]

Most participants shared the view that the CAs who used the disclaimer reminded them not to automatically follow the advice without question. A participant echoed this sentiment:

> *It's [CA with disclaimer] like okay, do your homework, you do more research. Don't just, like, take my word as Gospel.* [P19]

The participants appreciated the reminder that the CA is not a replacement for the advice of a medical practitioner:

> *I think it's responsible that they put it there. Because sometimes people jump to conclusions and then they get even more sick because they don't really know what they are doing. It's important to have that [disclaimer]. When they [the CA] were like "consult your doctor," I was like okay well maybe I might need to.* [P19]

A participant expanded this view and explored how using a medical disclaimer may potentially help limit the spread of misinformation through its precautionary message:

> *I felt secured like...in this day of social media there's so much misinformation out there. I have some coworkers that believe in those...conspiracy theories. So like if other people were to use this...device people would take it to heart. So I think it's really important that we have...that disclaimers [are] added at the end.* [P29]

Throughout the interviews, the participants expressed that the medical disclaimers increased their confidence in the CA, their sense of safety, their trust that the CA was a viable medication assistant, and influenced them to reconsider the CA's incorrect advice. However, these feelings were not shared by all participants. Our findings further reveal nuances reflected in the data with regard to the CAs' use of medical disclaimers and warnings.

Disclaimers: Creating Confusion and Redundancy

For some, the use of the verbal disclaimer was superfluous. A participant described the warning as redundant:

> *I felt like it [the disclaimer] was kind of stating the obvious to be honest...you would probably expect that that's not coming from a medical professional. So if you are asking that question, you're kind of already accepting that.* [P26]

This comment demonstrated that, for some users, disclaimers do not communicate novel information but instead what they perceive as obvious information: the CA is not a clinician. In addition, although we incorporated the disclaimer as a risk communication strategy to increase user safety, some users expressed that the inclusion of the disclaimer ultimately communicated that the designers were concerned with avoiding potential legal liability:

> *I think the disclaimer is just...CY. Cover Yourself. You have to say that.* [P23]

Such a viewpoint can ultimately diminish a user's perception of risk as well as the effectiveness of the precautionary warning and negatively affect the user's trust in the device.

Other participants pointed out that the warning appended to the advice increased the length of the auditory information considerably. A participant explored these drawbacks when she stated as follows:

> *But in terms of how I process the information, um, that it was giving me, it just added on to the amount of information, and it kind of made it more confusing.* [P21]

Several participants agreed that when using a conversational system that relies on processing and understanding auditory information alone, the disclaimer obfuscated the CA's answer, making the exchange inefficient.

Beyond inefficiency, the participants also expressed sensitivity toward warning fatigue. They stated that if a disclaimer were used during every interaction, they would stop taking the warning seriously and discontinue their use of the CA entirely, reflecting prior findings on speech-based warnings [38]. A participant succinctly stated as follows:

> *[The disclaimer] makes me feel like why am I wasting my time with this [CA] when I should be going to a real professional?* [P29]

By reminding the user that the system was not a replacement for the advice of a medical practitioner, some participants not only reconsidered the accuracy of the advice, but they also determined that the CA's functionality as a medication assistant was limited.

Paraphrasing: Developing Trust and Facilitating Confidence

When describing the properties related to making an informed medical decision, the participants explored how trusting the information source is critical. They spoke of trusting the CA that demonstrated conversational understanding. A participant described how grounding affected her perceptions of the system:

> *She [good paraphrase CA]...was geared exactly to what I was asking, um, and yeah...she just gave me the most confidence in, in the answer that I received.* [P19]

The participants compared CAs that accurately or inaccurately paraphrased their questions. They reported that an inaccurate paraphrase decreased the likelihood that they would follow the CA's advice. A participant stated as follows:

> *She [bad paraphrase CA] just didn't really understand what I was asking, so I felt uneasy about the information.* [P20]

Mistakes such as a bad paraphrase at the beginning of the interaction lowered users' confidence in the CA's abilities, causing users to immediately question the soundness of the CA's advice. A participant described how quickly a bad paraphrase creates doubt:

> *But for the bad paraphrasing, like right away when talking, you just know that they're providing me wrong information right away.* [P29]

However, when the CA paraphrased the participant's query correctly, the participants reported higher confidence in the accuracy of the CA's incorrect advice. A participant emphatically stated as follows:

> *Oh, absolutely, [good paraphrase] is number one. The trust comes right there. I said "this." You [the CA] listened.* [P23]

From the participants' perspective, the CA's use of a good paraphrase not only demonstrated a certain level of conversational understanding, but was also perceived as a meaningful and responsive component of the conversational exchange. In comparison, the disclaimer was described as a tacked-on *canned* statement. From the participants' perspective, the disclaimer would be present in the conversation irrespective of what the participant actually said, and the disclaimer's overall contribution to the system was mainly as a functional safety alert.

Discussion

Principal Findings

We found that grounding the feedback provided by a CA, in the form of paraphrases of user input, was effective at decreasing potentially harmful actions by the participants when the feedback indicated that the CA did not fully understand their query, supporting H1. We also found that signaling a lack of understanding significantly increased the likelihood that the participants would seek additional information before acting on the CA's recommendations. A warning message that the CA's advice should not be taken as an alternative to medical advice from a physician was effective at increasing the likelihood that the participants would consult a physician before acting on the CA's advice, supporting H2.

When interviewed, several of the participants indicated that disclaimers had benefits beyond promoting safe behavior, for example, by increasing the credibility of the device and their sense of reassurance and security in using it. However, several participants also indicated that disclaimers should be used sparingly and kept as brief as possible to avoid obfuscating the CA's response by adding content in the limited audio channel.

Grounding, in the form of paraphrasing participant queries, was cited as being important in assessments of credibility and trust, at least when the grounding indicated that the CA had properly understood a query. Incorrect paraphrases not only led to a decreased likelihood of acting on the CA's advice, but also affected negatively the participant's assessment of the CA and desire to use it in the future.

Our quantitative and qualitative findings demonstrate that accurate grounding increased the participants' confidence in the CA's medical advice by signaling that the user was properly heard. In this experiment, all our CAs relayed harmful medication advice. As a result, grounding alone was insufficient for mitigating user risk and potentially could have misled the participants to act on harmful medication advice. A participant described their response to a good paraphrase CA that did not incorporate a disclaimer as follows:

> *I felt like she did understand the full scope of the question and then was subsequently able to answer it by saying that it was safe to take.* [P32]

When asked directly for their perceptions of the CA that used a disclaimer, this participant explored how the addition of a disclaimer could keep users safe:

> *I would say…[the disclaimer] can reinforce that if you're not entirely sure or in the instance where maybe…you are impulsively doing something, that reinforcement that maybe you do need to seek another opinion could sway you from doing something that maybe you should or shouldn't do.* [P32]

Limitations

Our study includes several limitations, including the small convenience sample used. Restricting eligibility to native speakers of English certainly skewed our sample, but based on pilot testing, CA sessions with nonnative speakers yielded insufficient data, given the extremely high nonrecognition rates. Limiting participants to scripted utterances decreases the validity of our findings compared with allowing them to query CAs in their own words. However, we were primarily interested in investigating user perceptions of mitigation strategies and feel that our controlled examples achieved that by using actual unconstrained participant queries and actual CA responses from a prior study. We did not assess the participants' prior knowledge of the specific medical topics that we used as examples in our study, and this could have biased our findings. Finally, CA trust, credibility, and warning fatigue change over

time and must ultimately be assessed in a longitudinal context. For example, some researchers have found that a user's familiarity with a product significantly decreases their tendency to attend to warnings [42,43], indicating that warnings may lose their effectiveness over time with regular product use. Our study examines only first impressions of the mitigation strategies evaluated.

Conclusions

Designers of conversational systems should consider incorporating both warning messages and grounding techniques in response to user medical queries where harm could occur if consumers act on the advice, whether it is correct or not. To decrease alarm fatigue, warnings should be used sparingly and only when a CA determines that the user is trying to obtain actionable medical advice. In contrast, grounding feedback should always be provided because it has utility for all kinds of queries, both medical and nonmedical.

Note that use of these techniques does not guarantee safety: a CA may fully understand the user's query and provide grounding evidence of its understanding, but it may still retrieve incorrect advice or the user may misunderstand it [11]. In these cases, grounding may actually result in misplaced trust and increased user intent to act on potentially harmful advice. Ultimately, to maximize safety, grounding should convey the CA's understanding of what the user understands about the advice given, as well as the CA's understanding of what the user plans to do with the requested information.

The potential for AI systems to cause harm has long been recognized [44,45], but CAs that provide advice through unconstrained natural language represent one of the most challenging types of systems to ensure safety for. There is now increased interest in addressing issues of bias, safety, and validity in black box AI natural language processing systems [46]. However, recently proposed approaches that focus on describing appropriate contexts of use or the use of validation test suites [19] cannot possibly cover all cases that could lead to user harm, given the very large number of contextualized discourses that are possible. Despite the high error rates currently exhibited by CAs and with no clear approach to ensuring their safety, many experts feel that CAs will soon be able to provide reliable medical advice. A Delphi panel that comprised managers, physicians, researchers, and industry experts concluded that CAs will be able to provide "solid medical advice" within the next 5 years [47]. We feel that this projection is not based on an in-depth understanding of the issues, and that risk mitigation strategies such as those we have outlined here are needed until approaches to provably minimize the potential for CAs to give harmful advice are developed.

We reiterate the conclusions in the study by Bickmore et al [11] that unconstrained natural language input—typed text or speech—should not be used in CAs that are designed primarily to provide laypersons with medical advice. Such CAs have the potential to cause harm if users act on incorrect advice without first consulting a health care professional. Consumers lack mental models of CAs and cannot know the extent of the CAs' medical expertise or their linguistic capabilities and, even with improved grounding, may fail to recognize when the CA does not properly understand their communicative intent, fail to recognize when the CA has retrieved incorrect information, or fail to properly understand the CA's advice. The 2 mitigation strategies that we have explored in this work should only be used on CAs intended for other purposes (eg, general use, such as Siri or Alexa) when users naively ask for medical advice.

Future Work

The development of AI models that are explainable is an active area of research that is highly relevant to the implementation of safe CAs [48]. Indeed, explanation of how a CA understands a user can be seen as grounding, and the development of these methods targeting layperson understanding is an important direction of investigation. Given the complexity of the underlying AI models used in state-of-the-art CAs, additional media beyond the voice channel may be required to provide users with a fuller understanding of what is behind CA advice.

Identification of potentially unsafe user queries is a prerequisite for delivering targeted warning messages and also represents an important area of research. Such identification is nontrivial, given the many potential contexts of use, user states, and user intents, but is an important area of research in its own right. Persistent knowledge of users' medical condition, medications, and other electronic health record information could be critical in the medical domain.

The interaction of disclaimers and grounding strategies should be further explored. For example, the presence or absence of disclaimers may affect how users engage with grounding strategies.

Longitudinal studies of user interactions with CAs are particularly important to assess changes in attitudes toward a CA and changes in reactions to warnings over time.

Establishing the prevalence of actual harm from incorrect medical advice from a CA would be important to further motivate this area of research, requiring large-scale epidemiological surveys of patients, consumers, and medical professionals.

Finally, there may be many additional mitigation strategies that could be explored, such as having a CA engage users in dialog to understand why they are asking for medical advice and what they intend to do with the information provided before any advice is provided.

Acknowledgments

The authors thank Elise Chan for assisting with the human participants experiment.

Authors' Contributions

TB developed the study protocol, analyzed the results, and drafted the manuscript. SO developed the study protocol and materials, developed the simulated conversational assistants, conducted sessions with the participants, conducted the quantitative analyses, and contributed to the manuscript. TO developed the study protocol and materials, conducted sessions with the participants, conducted the qualitative analyses, and contributed to the manuscript.

Conflicts of Interest

None declared.

References

1. Schindler-Ruwisch J, Esposito CP. "Alexa, am I pregnant?": a content analysis of a virtual assistant's responses to prenatal health questions during the COVID-19 pandemic. Patient Educ Couns 2021 Mar;104(3):460-463 [FREE Full text] [doi: 10.1016/j.pec.2020.12.026] [Medline: 33422368]
2. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
3. Yang S, Lee J, Sezgin E, Bridge J, Lin S. Clinical advice by voice assistants on postpartum depression: cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e24045 [FREE Full text] [doi: 10.2196/24045] [Medline: 33427680]
4. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075 [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
5. Ferrand J, Hockensmith R, Houghton RF, Walsh-Buhi ER. Evaluating smart assistant responses for accuracy and misinformation regarding human papillomavirus vaccination: content analysis study. J Med Internet Res 2020 Aug 03;22(8):e19018 [FREE Full text] [doi: 10.2196/19018] [Medline: 32744508]
6. Boyd M, Wilson N. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS One 2018;13(3):e0194811 [FREE Full text] [doi: 10.1371/journal.pone.0194811] [Medline: 29590168]
7. Wilson N, MacDonald EJ, Mansoor OD, Morgan J. In bed with Siri and Google Assistant: a comparison of sexual health advice. Br Med J 2017 Dec 13;359:j5635. [doi: 10.1136/bmj.j5635] [Medline: 29237603]
8. Nobles AL, Leas EC, Caputi TL, Zhu S, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. NPJ Digit Med 2020;3:11 [FREE Full text] [doi: 10.1038/s41746-019-0215-9] [Medline: 32025572]
9. Picard C, Smith KE, Picard K, Douma MJ. Can Alexa, Cortana, Google Assistant and Siri save your life? A mixed-methods analysis of virtual digital assistants and their responses to first aid and basic life support queries. BMJ Innov 2020 Jan 07;6(1):26-31. [doi: 10.1136/bmjinnov-2018-000326]
10. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
11. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
12. Kinsella B. Voice assistant use on smartphones rise, siri maintains top spot for total users in the U.S. Voicebot. 2020 Nov 05. URL: https://voicebot.ai/2020/11/05/voice-assistant-use-on-smartphones-rise-siri-maintains-top-spot-for-total-users-in-the-u-s/ [accessed 2021-09-30]
13. Kinsella B. Nearly 90 million US Adults have smart speakers, adoption now exceeds one-third of consumers. Voicebot. URL: https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/ [accessed 2021-09-30]
14. Georgiou M. Why (and how) you should optimize your website for voice search. Entrepreneur. 2019 Dec 05. URL: https://www.entrepreneur.com/article/343036 [accessed 2021-09-30]
15. Bentley F, Luvogt C, Silverman M, Wirasinghe R, White B, Lottridge D. Understanding the long-term use of smart speaker assistants. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018 Sep Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 2018; - p. 1-24. [doi: 10.1145/3264901]
16. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. In: Proceedings of the ACM Transactions on Computer-Human Interaction. 2020 Sep 25 Presented at: ACM Transactions on Computer-Human Interaction; September 2020 p. 1-27. [doi: 10.1145/3373759]

17. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
18. Bujnowska-Fedak MM, Węgierek P. The impact of online health information on patient health behaviours and making decisions concerning health. Int J Environ Res Public Health 2020 Jan 31;17(3):880 [FREE Full text] [doi: 10.3390/ijerph17030880] [Medline: 32023828]
19. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Jan 29 Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; January 29 - 31, 2019; Atlanta, GA, USA p. 220-229. [doi: 10.1145/3287560.3287596]
20. Miner AS, Haque A, Fries JA, Fleming SL, Wilfley DE, Wilson GT, et al. Assessing the accuracy of automatic speech recognition for psychotherapy. NPJ Digit Med 2020;3:82 [FREE Full text] [doi: 10.1038/s41746-020-0285-8] [Medline: 32550644]
21. Brundage M, Avin S, Wang J, Belfield H, Krueger G, Hadfield G, et al. Toward trustworthy AI development:mechanisms for supporting verifiable claims. arXiv. 2020. URL: https://arxiv.org/pdf/2004.07213.pdf [accessed 2021-09-30]
22. Ziff P. The number of english sentences. Found Lang 1994 Jul 19;11(4):519-532 [FREE Full text]
23. Grosz B, Pollack M, Sidner C. Discourse. In: Posner MI, editor. Foundations of Cognitive Science. Cambridge, MA: MIT Press; 1989.
24. Clark H, Brennan S. Grounding in communication. In: Resnick LB, Levine JM, Teasley SD, editors. Perspectives on Socially Shared Cognition. Washington: American Psychological Association; 1991.
25. Clark H, Schaefer E. Contributing to discourse. Cogn Sci 1989 Apr;13(2):259-294. [doi: 10.1207/s15516709cog1302_7]
26. Grice HP. Logic conversation. In: Cole P, Morgan JL, editors. Syntax and Semantics. New York: Academic Press; 1975:41-58.
27. Bohus D, Rudnicky AI. Sorry, I didn't catch that! An investigation of non-understanding errors and recovery strategies. In: Dybkjaer L, Minker W, editors. Recent Trends in Discourse and Dialogue. Dordrecht: Springer; 2008:123-154.
28. Skantze G. Exploring human error recovery strategies: implications for spoken dialogue systems. Speech Commun 2005 Mar;45(3):325-341. [doi: 10.1016/j.specom.2004.11.005]
29. Cho J, Rader E. The role of conversational grounding in supporting symbiosis between people and digital assistants. In: Proceedings of the ACM on Human-Computer Interaction. 2020 May Presented at: ACM on Human-Computer Interaction; May 2020; - p. 1-28. [doi: 10.1145/3392838]
30. Yuan S, Bruggemeier B, Hillmann S, Michael T. User preference and categories for error responses in conversational user interfaces. In: Proceedings of the 2nd Conference on Conversational User Interfaces. 2020 Jul 22 Presented at: CUI '20: 2nd Conference on Conversational User Interfaces; July 22 - 24, 2020; Bilbao Spain p. 1-8. [doi: 10.1145/3405755.3406126]
31. Yaghoubzadeh R, Pitsch K, Kopp S. Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users. In: Proceedings of the International Conference on Intelligent Virtual Agents (IVA). 2015 Aug 01 Presented at: International Conference on Intelligent Virtual Agents (IVA); August 26-28, 2015; Delft, The Netherlands p. 22-38. [doi: 10.1007/978-3-319-21996-7_3]
32. Aneja D, McDuff D, Czerwinski M. Conversational error analysis in human-agent interaction. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. 2020 Oct 20 Presented at: IVA '20: ACM International Conference on Intelligent Virtual Agents; October 20 - 22, 2020; Virtual Event Scotland UK p. 1-8. [doi: 10.1145/3383652.3423901]
33. Prakash A, Gupta V. Would you trust a bot for healthcare advice? An empirical investigation. In: Proceedings of the 24th Pacific Asia Conference on Information Systems: Information Systems (IS) for the Future, PACIS 2020. 2020 Jun Presented at: 24th Pacific Asia Conference on Information Systems, PACIS 2020; June 20 - 24, 2020; Dubai, UAE.
34. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]
35. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manage Inf Syst 2011 Jun 11;2(2):1-25. [doi: 10.1145/1985347.1985353]
36. Corneliussen SE. Factors affecting intention to use chatbots for health information : a qualitative exploratory study. Master thesis at the Department of Psychology, University of Oslo. 2020. URL: https://www.duo.uio.no/bitstream/handle/10852/79932/Susanna-CorneliussenV2020.pdf?sequence=1&isAllowed=y [accessed 2021-10-13]
37. Ruiter RA, Abraham C, Kok G. Scary warnings and rational precautions: a review of the psychology of fear appeals. Psychol Health 2001 Nov;16(6):613-630. [doi: 10.1080/08870440108405863]
38. Noyes JM, Hellier E, Edworthy J. Speech warnings: a review. Theor Issues Ergon Sci 2006 Nov;7(6):551-571. [doi: 10.1080/14639220600731123]
39. Davis T, Long S, Jackson R, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
40. Medical disclaimer. SEQ Legal. URL: https://seqlegal.com/free-legal-documents/medical-disclaimer [accessed 2021-09-30]
41. Schwartz LM, Woloshin S. Medical marketing in the United States, 1997-2016. J Am Med Assoc 2019 Jan 01;321(1):80-96. [doi: 10.1001/jama.2018.19320] [Medline: 30620375]

42. Godfrey SS, Allender L, Laughery KR, Smith VL. Warning messages: will the consumer bother to look? Proc Hum Factors Ergon Soc Annu Meet 1983 Oct 01;27(11):950-954. [doi: 10.1177/154193128302701118]
43. Wogalter MS, Brelsford JW, Desaulniers DR, Laughery KR. Consumer product warnings: the role of hazard perception. J Safety Res 1991;22(2):71-82. [doi: 10.1016/0022-4375(91)90015-n]
44. Asimov I. I, Robot. London: Harper Voyager; 2013.
45. Asilomar AI principles. Future of Life Institute. 2017. URL: https://futureoflife.org/ai-principles/ [accessed 2021-09-30]
46. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. J Am Med Assoc 2019 Dec 24;322(24):2377-2378. [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
47. Ermolina A, Tiberius V. Voice-controlled intelligent personal assistants in health care: International Delphi Study. J Med Internet Res 2021 Apr 09;23(4):e25312 [FREE Full text] [doi: 10.2196/25312] [Medline: 33835032]
48. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020 Jun;58:82-115. [doi: 10.1016/j.inffus.2019.12.012]

Abbreviations

AI: artificial intelligence
CA: conversational assistant
ECA: embodied conversational agent
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 26.05.21; peer-reviewed by V Tiberius, E Grekin; comments to author 17.08.21; revised version received 29.08.21; accepted 18.09.21; published 09.11.21.

Please cite as:
Bickmore TW, Ólafsson S, O'Leary TK
Mitigating Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: Exploratory Mixed Methods Experiment
J Med Internet Res 2021;23(11):e30704
URL: https://www.jmir.org/2021/11/e30704
doi:10.2196/30704
PMID:34751661

Original Paper

Patient Interactions With an Automated Conversational Agent Delivering Pretest Genetics Education: Descriptive Study

Daniel Chavez-Yenter[1,2], BS, MPH; Kadyn E Kimball[2], BS; Wendy Kohlmann[3], MS, CGC; Rachelle Lorenz Chambers[4], MS; Richard L Bradshaw[5], MS, PhD; Whitney F Espinel[3], MMSc; Michael Flynn[6], MD; Amanda Gammon[3], MS, CGC; Eric Goldberg[7], MD; Kelsi J Hagerty[3], BS; Rachel Hess[8], MS, MD; Cecilia Kessler[3], BS; Rachel Monahan[4,9], BA; Danielle Temares[4], BS; Katie Tobik[3], MGCS; Devin M Mann[9], MS, MD; Kensaku Kawamoto[5], MHS, MD, PhD; Guilherme Del Fiol[5], MD, PhD; Saundra S Buys[3,10], MD; Ophira Ginsburg[4,9], MSc, MD; Kimberly A Kaphingst[1,2], ScD

[1]Department of Communication, University of Utah, Salt Lake City, UT, United States

[2]Cancer Control and Population Sciences, Huntsman Cancer Institute, Salt Lake City, UT, United States

[3]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States

[4]Perlmutter Cancer Center, New York University Langone Health, New York, NY, United States

[5]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

[6]University of Utah Health, Salt Lake City, UT, United States

[7]Department of Medicine, New York University Grossman School of Medicine, New York University, New York, NY, United States

[8]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States

[9]Department of Population Health, New York University Grossman School of Medicine, New York University, New York, NY, United States

[10]Department of Internal Medicine, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Daniel Chavez-Yenter, BS, MPH
Department of Communication
University of Utah
255 S Central Campus Drive
Salt Lake City, UT, 84112
United States
Phone: 1 801 213 5724
Email: daniel.chavez-yenter@utah.edu

Abstract

Background: Cancer genetic testing to assess an individual's cancer risk and to enable genomics-informed cancer treatment has grown exponentially in the past decade. Because of this continued growth and a shortage of health care workers, there is a need for automated strategies that provide high-quality genetics services to patients to reduce the clinical demand for genetics providers. Conversational agents have shown promise in managing mental health, pain, and other chronic conditions and are increasingly being used in cancer genetic services. However, research on how patients interact with these agents to satisfy their information needs is limited.

Objective: Our primary aim is to assess user interactions with a conversational agent for pretest genetics education.

Methods: We conducted a feasibility study of user interactions with a conversational agent who delivers pretest genetics education to primary care patients without cancer who are eligible for cancer genetic evaluation. The conversational agent provided scripted content similar to that delivered in a pretest genetic counseling visit for cancer genetic testing. Outside of a core set of information delivered to all patients, users were able to navigate within the chat to request additional content in their areas of interest. An artificial intelligence–based preprogrammed library was also established to allow users to ask open-ended questions to the conversational agent. Transcripts of the interactions were recorded. Here, we describe the information selected, time spent to complete the chat, and use of the open-ended question feature. Descriptive statistics were used for quantitative measures, and thematic analyses were used for qualitative responses.

Results: We invited 103 patients to participate, of which 88.3% (91/103) were offered access to the conversational agent, 39% (36/91) started the chat, and 32% (30/91) completed the chat. Most users who completed the chat indicated that they wanted to

continue with genetic testing (21/30, 70%), few were unsure (9/30, 30%), and no patient declined to move forward with testing. Those who decided to test spent an average of 10 (SD 2.57) minutes on the chat, selected an average of 1.87 (SD 1.2) additional pieces of information, and generally did not ask open-ended questions. Those who were unsure spent 4 more minutes on average (mean 14.1, SD 7.41; *P*=.03) on the chat, selected an average of 3.67 (SD 2.9) additional pieces of information, and asked at least one open-ended question.

Conclusions: The pretest chat provided enough information for most patients to decide on cancer genetic testing, as indicated by the small number of open-ended questions. A subset of participants were still unsure about receiving genetic testing and may require additional education or interpersonal support before making a testing decision. Conversational agents have the potential to become a scalable alternative for pretest genetics education, reducing the clinical demand on genetics providers.

(J Med Internet Res 2021;23(11):e29447) doi:10.2196/29447

KEYWORDS

cancer; genetic testing; virtual conversational agent; user interaction; smartphone; mobile phone

Introduction

Background

Cancer genetic testing and the use of genomic information are central to the future of precision cancer medicine [1,2]. Effective communication of germline genetic results and family history of cancer to patients is key to their understanding of their own (and in many cases their family members') cancer risks, evidence-based options for decision-making (ie, for cancer risk management), and in the case of individuals with cancer, their overall cancer trajectory, including treatment options [3]. Automated approaches to communication have begun to emerge as a way of meeting the expanding volume of testing in the context of a limited number of genetic counselors able to provide services [4]. One such approach is the use of an automated conversational agent to deliver cancer genetic services to supplement, or in lieu of, a genetic counselor. Conversational agents are automated, scripted, and responsive agents used to mimic human interactions. These agents use natural language processing to analyze user inputs and respond appropriately using human language via auditory or textual methods [4]. Conversational agents are increasingly popular in various health contexts, as they can be easily accessed through smartphones, tablets, laptops, or desktop computers. The agents are fairly accessible to most adults in the United States, of whom 75% report having at least one smartphone [5]. With an exponential growth in genetic testing as a way to identify individuals with inherited cancer susceptibility, conversational agents present an innovative approach to broadening access to clinical cancer genetics services in the face of limited health professionals with genomics expertise while encouraging wider use of cancer genetic testing incorporated in health care settings [6,7].

The delivery of health services through conversational agents in research contexts has been successfully tested in various health domains, such as mental health, asthma, diabetes management, and physical activity uptake [8]. Conversational agents have been found to help health care providers lower the rates of depression and anxiety [9-11] and also improve adherence to treatment for asthma, diabetes, and pain [12-14]. Recent research has begun to use the conversational agent model of care to facilitate informed decision-making and technology use self-efficacy related to prostate cancer [15,16]. Other conversational agents in noncancer settings, specifically mental health and lifestyle change interventions, identified in the Bibault et al [15] review, were found to be able to improve decision-making processes. Owens et al [16] developed iDecide, an embodied conversational agent–led, computer-based prostate cancer screening decision aid. Their findings showed that conversational agents were able to improve prostate cancer knowledge and informed decision-making self-efficacy and technology use self-efficacy among their target audience of African American men. Prior research on conversational agents has shown that most users are receptive to the use of this technology in health, although concerns related to accuracy, security, and lack of empathy have been raised [17,18]. Although conversational agents are a promising technology in various health contexts, more research is needed to examine their efficacy and implementation effectiveness and outcomes in genetic service delivery [19].

Conversational agents can be used in various ways in the context of cancer genetic services, especially in the context of hereditary cancers, for uses such as collecting initial data and communicating risk. Hereditary factors can affect the risk for many common adult-onset cancers (ie, breast, ovarian, colorectal, pancreas, and prostate) [20-26]. Cancer genetic testing generates quantifiable cancer risks, which can be helpful in directing the clinical management of patients. Such genetic risk assessments involve the collection of detailed patient information, such as their family's cancer history; delivery of pretest genetic counseling to inform decisions about testing; and returning the results after testing. Only recently have clinical cancer care settings begun to leverage this technology to support patient information management in service delivery [4]. Conversational agents have been used in some of these processes, such as collecting patient data, providing genetic information, delivering results, and facilitating *cascade testing* of at-risk relatives in clinical settings [18,27-30].

There has been limited research investigating how users interact with conversational agents in various contexts, including health, education, and customer service [31,32]. For this study, interactions will be characterized as users' reciprocal actions with our automated conversational agent. More research is needed on how conversational agents are used in health contexts, such as cancer genetic services and user interactions. For example, the Geisinger health care system has integrated a personalized conversational agent into the delivery of genetic

testing services. This agent is involved in obtaining consent, facilitating family sharing opportunities, and providing a return of results [18]. The acceptability of this approach in health care systems has been shown through qualitative studies [16-18]. In other words, through individual interviews and focus groups, users were asked about their experience using the personalized conversational agent and whether they found it an acceptable alternative to talking with an actual genetic counselor in regard to obtaining consent, cascade testing, and returning of results.

Objective

Despite the increasing use of conversational agents clinically and their initial acceptance, studies have yet to examine how patients interact with conversational agents implemented within a health care system to deliver cancer genetic services. To address this important research need, we will report on user interactions with a conversational agent in the delivery of pretest genetics education in a feasibility study through descriptive analysis of the information selected, time spent to complete the chat, and use of the open-ended question feature.

Methods

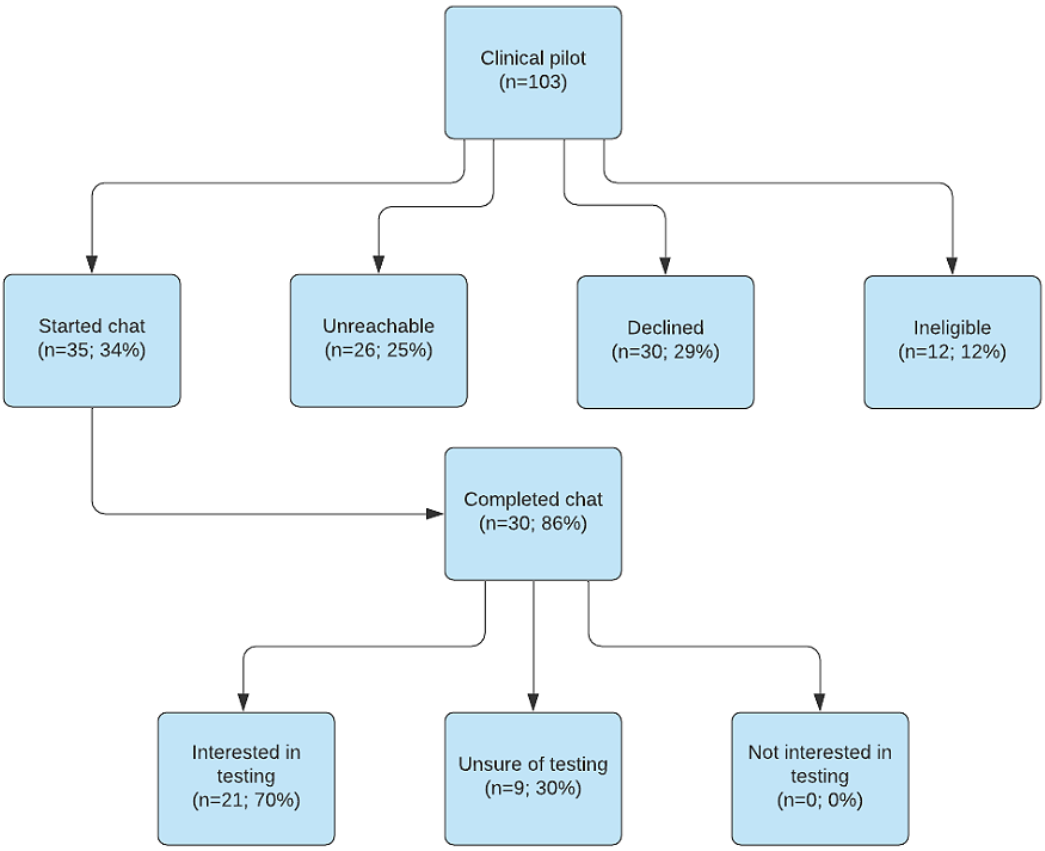
Participants

For this study, our team used a standards-based clinical decision support infrastructure to identify primary care patients without cancer in the University of Utah's health care system who were eligible for cancer genetic evaluation. This algorithm used cancer family history data available in the electronic health record (EHR) to identify those who met the National Comprehensive Cancer Network guidelines for genetic testing for hereditary breast or ovarian [33] and colorectal cancer based on their family history [34]. All identified patients in this study were English-speaking, between the ages of 25 and 60 years, had a primary care appointment in the past 3 years, had no prior cancer diagnosis other than nonmelanoma skin cancer, had no prior genetic counseling or testing related to hereditary cancer, and had a patient portal account.

Study Procedures

From February to June 2020, a sample of 103 identified patients received a message through the patient portal about their eligibility for genetic services and an invitation to complete a pretest genetics education chat with the conversational agent (Figure 1). Of the 103 patients, 12 (11.7%) were ineligible for the study (because of relocation, previous testing, or incomplete family history) as determined through an EHR review and follow-up communication with the patient. The patients who did not complete the chat received a second patient portal message and up to three follow-up telephone calls to encourage engagement. Once the chat was completed, patients were contacted by a genetic counseling assistant who could answer questions and facilitate genetic testing for those who opted to test. A transcript of the pretest genetics education chat was added to the patient's EHR and used for the analysis of interactions in this study. The study was approved by the University of Utah institutional review board.

Figure 1. Study flow diagram.



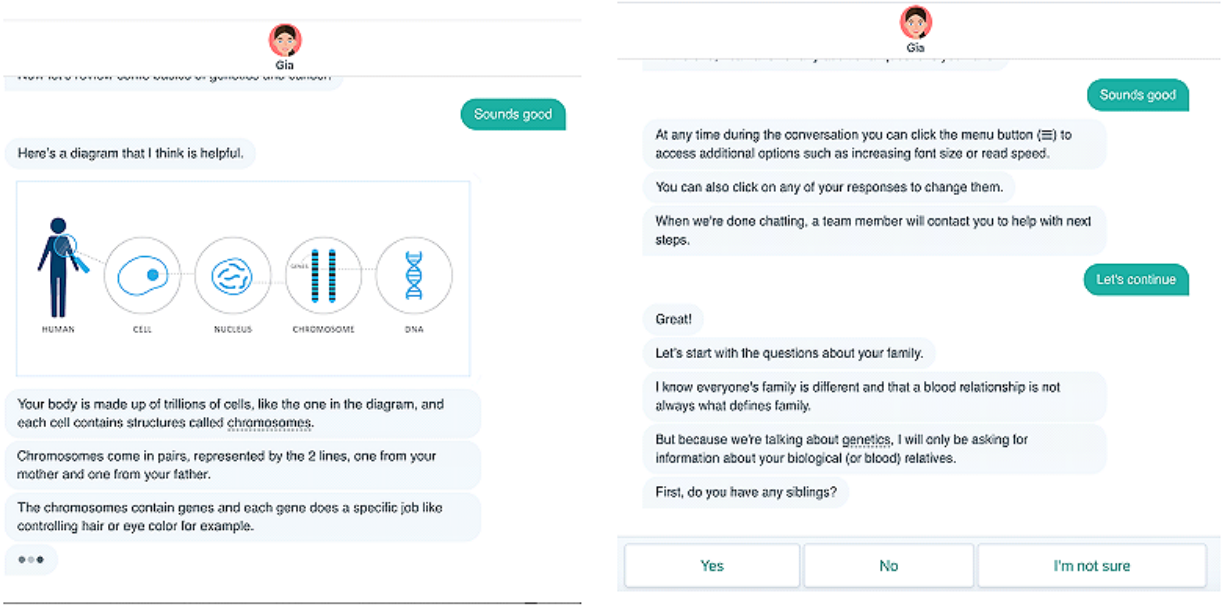
Content of Pretest Genetics Education Chat

The chat was developed by using the Health Insurance Portability and Accountability Act compliant with the Invitae platform. The content of the chat was based on the content of pretest genetic counseling delivered by certified genetic counselors at the Huntsman Cancer Institute, University of Utah, and the Perlmutter Cancer Center, New York University (Multimedia Appendix 1). Pretest genetic counseling provides information related to the purposes of genetic testing, genetic testing options, and the possible results generated [34-36]. Using an iterative design process, our interdisciplinary team, which comprised experts in genetic counseling, health communication, primary care, and cancer clinical genetics, scripted the chat content based on recordings of genetic counseling visits and clinical experience of the genetic counselors. Two rounds of both qualitative and quantitative user testing were completed to refine the content before the feasibility study described here.

The scripted pretest genetics education chat covered the major content areas in the following order: heritability of the risk for cancer, cancer risks associated with a mutation, genetic testing process, description of the types of genetic tests, and possible genetic testing outcomes and their implications. At the end of the chat with the conversational agent, patients were asked whether they were interested in continuing with the testing (yes, no, or unsure).

When patients opened the chat, the conversational agent first oriented them to the response buttons, menu (eg, to change text size and speed), and purpose of the chat. In addition, there was an introductory video filmed with the lead genetic counselor at the Huntsman Cancer Institute explaining the purpose of the chat and the conversational agent in the context of cancer genetic testing to add a human face and gain credibility with the user. The conversational agent then provided the scripted information in a real-time message format, with text bubbles containing 3 periods (ellipsis) to mirror an instant message conversation (Figure 2). All patients saw a core set of content decided upon by the team as essential for pretest genetics education. To supplement this core content at predetermined points throughout the conversation, options for responses were given on the bottom of the screen, and the patient's response determined the next content delivered. Additional information included a range of topics, such as asking for basic explanations of cancer and genetics, the various possible results of genetic testing, and genetic mutation risk for various cancer types (eg, breast, pancreatic, and colon). This process allowed patients to choose to receive more or less details on a key topic (eg, goal, benefit, and risk) based on their preferences.

Figure 2. Genetic information assistant screenshots.



In addition, throughout the chat, participants were able to ask free-text questions. Natural language processing was used to answer the questions, if possible, from a library of prescripted responses. This created a real-time communication experience with immediately answered inquires that did not require additional effort from a provider for a response. If the platform's artificial intelligence (AI) library was unable to match an answer to the patient's open-ended question, they were prompted with alternative questions related to the topic of interest that had prescripted answers assigned. If the system was not able to determine an appropriate response to a question, it was sent via email to the clinical care team.

Analysis

From the EHR, we abstracted data on patients' sex, age, family history of cancer, race, and ethnicity. To analyze how patients interacted with the conversational agent, we collected the following information from the chat transcripts upon completion of the chat: time spent interacting with the conversational agent (with the exclusion of 2 cases in which users left the chat idle

for >12 hours on their browser, for which we were unable to ascertain full time in interaction), options selected within the chat for supplemental content, and the number and types of open-ended questions asked by patients. A total of 2 coders independently extracted transcript data for 20% of the transcripts and had 94%-100% agreement for each code. Discrepancies were resolved by the full research team. From the study transcript records, we abstracted whether patients decided to continue with genetic testing after the chat. Data analysis was completed using SPSS version 25 (IBM Corp). We conducted 2-tailed *t* tests to examine differences in continuous variables between those patients who were sure of pursuing genetic testing and those who were not sure, as well as in demographics. Statistical significance was determined as $P<.05$.

Results

Overview

We identified and sent outreach messages via the patient portal to 103 patients. Of those 103, 12 (11.7%) were ineligible for the study (because of relocation, previous testing, or incomplete family history) and thus were excluded from the rest of the analysis. Of the 91 eligible patients, 75 (82%) opened the patient portal message with the chat link. About half of those participants (36/75, 48%) clicked the link to start the chat. Of those who started the chat, most finished it (30/36, 83%). As shown in Table 1, most of the patients who completed the chat were female (23/30, 77%), White (28/30, 93%), non-Hispanic or Latino (27/30, 90%), and had a mean age of 43.3 years (SD 9.96 years). Most patients had family histories of breast, ovarian, and pancreatic cancers, with no significant differences in these characteristics between those who did and did not complete the chat.

Table 1. Demographics and family histories of all invited users (N=91).

Demographics	Completed chat (n=30)	Did not complete chat (n=61)
Age (years), mean (SD)	43.3 (10.0)	42.1 (10.2)
Sex, n (%)		
Female	23 (77)	50 (82)
Male	7 (23)	10 (16)
Race, n (%)		
White	28 (93)	55 (90)
Black	N/A[a]	1 (2)
Asian	N/A	2 (3.2)
Other or did not disclose	2 (7)	2 (3.2)
Ethnicity, n (%)		
Non-Hispanic or Latino	27 (90)	55 (90)
Hispanic or Latino	2 (7)	4 (6)
Did not disclose	1 (3)	1 (2)
Family history of cancer,[b] n (%)		
Breast	10 (30)	20 (28)
Ovarian	9 (27)	21 (30)
Pancreatic	8 (24)	19 (27)
Colon	4 (12)	6 (8)
Prostate	1 (3)	4 (6)
Stomach	1 (3)	1 (1)

[a]N/A: not applicable.

[b]Patients could have >1 family member with a history of cancer, so the percentage values are not mutually exclusive.

At the completion of the educational content, users were asked if they wished to proceed with testing. As shown in Table 2, most users who completed the chat wished to continue with testing (21/30, 70%). Of these 30 users, 9 (30%) were unsure, whereas none indicated that they did not want to test. Age did not differ across groups (44 vs 42 years; $P=.66$); however, total time spent on the chat (10-14.1 minutes; $P=.03$), requests for additional information (1.2-4 requests; $P=.03$), and open-ended questions (0.3-1 open-ended questions; $P<.001$) did differ between those who decided to test and those who were unsure. The range of total time interacting with the chat was 6-31 minutes; however, most users spent 15-20 minutes with the chat. Those who indicated that they wanted to test spent an average of 10 (SD 2.57) minutes on the chat, selected 1-2 (mean

1.87, SD 1.2) options requesting additional pieces of information, and generally did not ask an open-ended question (Table 2). Those who were unsure spent 4 more minutes on average (mean 14.1, SD 7.41) with the chat and selected 3-4 (mean 3.67, SD 2.9) options requesting additional information. Of the 9 unsure patients, 4 (44%) asked 2 open-ended questions, with 1 participant asking 3 open-ended questions and 2 asking no open-ended questions.

Table 2. Completed chat continuous measures (N=30).

Continuous measures	Decided to test (n=21), mean (SD; range)	Unsure about testing (n=9), mean (SD; range)	All, mean (SD; range)
Age (years)	43.81 (9.15; 30-59)	42.0 (12.14; 25-60)	43.27 (9.96; 25-60)
Time spent on chat (minutes)[a,b]	10.0 (2.57; 6-15)	14.1 (7.41; 9-31)	11.17 (4.71; 6-31)
Total additional information items asked (GIA[c] initiated)[b]	1.87 (1.24; 0-5)	3.67 (2.92; 1-9)	2.4 (2.0; 0-9)
Total open-ended questions asked[b]	0.095 (0.30; 0-1)	1.11 (1.05; 0-3)	0.4 (0.77; 0-3)

[a]N=28; 2 cases were excluded because the total time spent on chat was not collected.

[b]$P<.05$ (exact P values reported in main text).

[c]GIA: genetic information assistant.

In examining patients' selections of options to request more information, there was a mean of 2.4 requests (SD 2.0; median 2.0, range 0-9) across all users. Only 1 patient asked for no additional pieces of information. As shown in Table 3, the most common topics for which patients requested more information were basic information about genetics and cancer (28/30, 93%), what types of risk factors were used to assess their risk (9/30, 30%), the genes that were included in the genetic test (9/30, 30%), and what options exist to lower cancer risk (9/30, 30%). Of the 10 patients who requested three or more additional pieces of information, all requested more information on basic genetics and cancer, 7 (70%) wanted more information on what genes were included in the test, and 8 (80%) wanted to know what options exist to mitigate cancer risk, whereas the rest of the selections varied.

Table 3. Use of options to request additional information (N=30).

Additional information requested	Value, n (%)
Basic information about cancer and genetics	28 (93)
Definition of sporadic causes of cancer	0 (0)
Increase in cancer risk related to a positive genetic test result	1 (3)
Reason for range of cancer risk estimates	0 (0)
Risk estimates for other cancers	0 (0)
How patients were identified	9 (30)
What genes are included on the test	9 (30)
Options for decreasing cancer risk	9 (30)
Surgical or medical options for decreasing cancer risk	1 (3)
Possible types of results from genetic testing	
Positive	3 (10)
Negative	2 (7)
Variant of unknown significance	3 (10)
Privacy protections	1 (3)
Any additional open-ended questions?	8 (27)
Participant initial decision	
Continue with testing	21 (70)
Unsure about continuing with testing	9 (30)

Open-ended Questions

About one-quarter of the patients (8/30, 27%) typed in an open-ended question; however, a total of 11 questions were posed to our AI-based preprogrammed library (Textbox 1). Of the 9 users, 5 (56%) asked 1 question, 3 (33%) asked 2 questions, and 1 (11%) asked 3 questions. Open-ended questions

often related to the cost of genetic testing with insurance and how this testing was different from direct-to-consumer genetic testing panels. Other question topics included how long testing results would take and whether health conditions could interfere with genetic testing results. Of the total 11 open-ended questions, 3 (27%) were directed to the clinical care team for follow-up and clarification; of these 3 questions, 1 (33%) was worded as a request rather than a question ("Let me know if I need to pay $0 out of pocket"). One open-ended question was clarified using the platform's AI library by providing alternate questions that related to the proposed open-ended question, with the patient selecting the closest one, whereas the remaining questions could be answered directly through scripted responses in the library.

Textbox 1. List of open-ended questions asked by patients.

Open-ended questions

- "Do you look at Aunts, Uncles, Cousins?"
- "Please let me know if my genetic testing would be greater than $0 out of pocket" (the questions were sent to the clinical care team).
- "How will this work with my insurance?"
- "Are there any health conditions that can interfere with the accuracy of the genetic test results?"
- "How much does the test cost?"
- "How long does it take to get test results?"
- "Is this different than 23andme and Ancestry?"
- "Why should I get screened?"
- "Will Healthy U Medcaid cover the cost of the test?"
- "How will this work with my insurance?"
- "What if I already have done the 'Color' genetic testing?"

Discussion

Principal Findings

Although conversational agents are increasingly important in precision medicine and are a potential alternative to educational sessions conducted by a person [11], there is a need to understand how users utilize and interact with the conversational agent to know what components (eg, topics and interface) may be most important to them. These data can add to prior findings on the acceptability and usability of conversational agents in multiple different contexts (behavior change, mental health, adherence, and genomics) [8-14,16-18]. Understanding what components are most salient and how users interact with the platform will enable providers, designers, and researchers to improve engagement with conversational agents for better patient experiences and potentially improved health outcomes. To our knowledge, this is the only study that has characterized the use of a conversational agent in a genomic setting, adding to the literature assessing its acceptability, usability, and understanding in various health contexts. Our findings suggest that the automated conversational agent approach engaged about one-third of the eligible patient population in clinical cancer genetic testing, with moderate outreach attempts and no health care appointments, highlighting potential scalability in its broader use as a potential cost-saving measure as well.

Although not directly asked of study participants, our conversational agent generally met the information needs of patients considering cancer genetic testing based on the limited use of open-ended questions. Patients asked for 2.4 additional pieces of information on average. More specifically, in addition to the core set of information that all patients received, patients mostly wanted more information on the basics of genetics and cancer, what types of risk factors were used to assess their risk, what genes were included in the genetic test, and what options exist to lower cancer risk. Additional information sought via open-ended questions related to cost, differences from direct-to-consumer genetic testing, whether health conditions could interfere with results, and how long results would take. A subset of patients spent more time on the chat, asked for additional pieces of information, asked more open-ended questions, and were unsure about testing at the end of the chat. Such high information–seeking patients may need additional support from a clinical provider to make a testing decision. However, the use of a conversational agent may substantially reduce provider burden by meeting the educational needs of most patients. Genetic counseling team members can then follow up with patients who have additional questions and concerns.

Our results are consistent with a previous study completed in the same patient population pool using the clinical decision support algorithm and standard of care approach involving a scheduled genetic counseling appointment [35]. Regardless of the delivery approach, both approaches (our automated conversational agent and the standard of care) reached just over a 30% unsolicited outreach uptake rate for clinical cancer genetic testing. It is important to note that testing use uptake did not differ by gender, age, and family history of cancer. However, it is important to consider other variables in the future, such as eHealth literacy and attitudes and beliefs related to cancer genetic testing uptake, as each approach may work differently and be most appropriate for different populations and within other contexts.

The findings further showed that when patients wanted additional information, they often selected one of the presented options rather than creating their own open-ended question, which highlights the importance for researchers, health care providers, and communication specialists to carefully design these options. Future research in genome-specific conversational agent contexts should examine why patients do and do not enter open-ended questions and whether there is a way to further improve chat interactivity via this feature. Our findings suggest that interactivity, in our case users interacting with our automated conversational agent, may be an important part of informing users regarding a particular topic, as suggested by the cognitive theory of multimedia learning [36] and adult learning theory [37]. Drawing more from these theories enables our understanding of how to leverage interactivity to encourage the elaboration of genetic testing technology while assessing knowledge increases. The development of conversational agents has underutilized theory and theory-driven concepts [8,38], such as interactivity. The greater use of theory may lead to more effective educational efforts as well as findings that are more generalizable across contexts, informing evidence-based strategies on how to best engage with users through interactions with conversational agents.

Our study expands upon a prior study that used focus groups to assess the acceptability, usability, and understanding of conversational agents for consent, follow-up, and cascade genomic testing [18]. The prior study found that users strongly supported the use of conversational agents in the context of providing genomic services. Although this prior study found that the use of the AI library of responses to open-ended questions was very appealing to users [18], in our study, we found that users made limited use of open-ended questions in actual practice. This could have also been because of the amount of predetermined information content in the conversational agent's script.

In one other use of a conversational agent for a related application in the context of chronic conditions (eg, cardiovascular disease, diabetes, and cancer), Wang et al [39] developed an animated virtual counselor to collect electronic family health histories for clinical risk assessment [29,30] and as a proxy for genetic predisposition to personalize medical care and disease prevention [40-45]. In a randomized comparison with the Surgeon General's My Family Health Portrait [46], the conversational agent had better acceptability and usability outcomes (eg, ease of use, flow, understanding information, and satisfaction). However, the study did not assess in detail how users interacted with the virtual counselor for knowing your family history. Therefore, we enhance the previous understanding of both the acceptability and usability of conversational agents and how their features are used in actual practice.

Strengths and Limitations

As our study was a feasibility study, it had some limitations. One of our limitations was the composition of our cohort, which primarily comprised White, older women. The use of conversational agents in other populations may differ. We restricted our study to patients aged 25-60 years because screening and prevention recommendations can be modified for those in this range with inherited cancer susceptibility or familial risk [47-49]. Our cohort being predominantly middle-aged may have affected the level of engagement we saw with the conversational agent. Second, we were not able to examine how long a patient actually interacted with the conversational agent as opposed to merely having the chat open. It is possible that users left the chat idle and came back to it at a later time. Eye-tracking and other laboratory-based studies will allow further examination of these issues. Finally, we did not directly ask the users if their informational needs were met, which is an important next step. However, with the lack of open-ended questions being asked by users, we believe that the conversational agent was able to effectively facilitate decision-making for cancer genetic testing. More research on user experience and participant perceptions of informational needs is needed in this area.

Conclusions

Despite these limitations, our study's results suggest that a conversational agent can meet the information needs of primary care patients and can represent a scalable alternative for pretest counseling for patients considering cancer genetic testing. With the increased demand for genetic testing and counseling, through the development, implementation, and maintenance of conversational agents, such as the one presented in our study, this strategy has the potential to save operating costs and improve the availability of these technologies for underserved groups. This indicates that our conversational agent may be an acceptable alternative (or supplement) to an in-person genetic counseling pretest visit, although outcomes such as use of testing should be evaluated in a randomized trial. In addition, we learned how patients interact with our conversational agent, what additional information is of most interest, and the patients' interest in using the open-ended question feature. We also found that patients who were unsure about testing tended to ask for more information, asked open-ended questions, spent more time with the chat, and may need additional interpersonal support and information for decision making.

Acknowledgments

The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health, United States, under award numbers U01CA232826 and U24CA204800. The funders had no role in the data collection, analysis or interpretation, writing of the report, or the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors would like to thank Kelsey Kehoe for her early contributions to collecting and coding data.

Conflicts of Interest

KK reports honoraria, consulting, sponsored research, licensing, or codevelopment outside the submitted work in the past 5 years with McKesson InterQual, Hitachi, Pfizer, Premier, Klesis Healthcare, Research Triangle Institute International, Mayo Clinic, the University of Washington, the University of California at San Francisco, Vanderbilt University, Medical Doctor Aware, and the United States Office of the National Coordinator for Health Information Technology (via Enterprising Science and Computing, security risk solutions, A+ government solutions, and Hausam Consulting) in the area of health information technology. KK was also an unpaid board member of the nonprofit Health Level Seven International health information technology standard development organization; he is an unpaid member of the United States Health Information Technology Advisory Committee, and he has helped develop a number of health information technology tools that may be commercialized to enable wider impact. None of these relationships have direct relevance to the manuscript but have been reported in the interest of full disclosure.

Multimedia Appendix 1
Genetic information assistant chat script for pretest educational conversation.
[DOCX File , 113 KB - jmir_v23i11e29447_app1.docx]

References

1. Yang YB. The role of genetics in medicine: a future of precision medicine. BC Medical Journal 2019;61(10):388-389 [FREE Full text]
2. Kensler TW, Spira A, Garber JE, Szabo E, Lee JJ, Dong Z, et al. Transforming cancer prevention through precision medicine and immune-oncology. Cancer Prev Res 2016 Jan 06;9(1):2-10. [doi: 10.1158/1940-6207.capr-15-0406]
3. Kaphingst KA, Peterson E, Zhao J, Gaysynsky A, Elrick A, Hong SJ, et al. Cancer communication research in the era of genomics and precision medicine: a scoping review. Genet Med 2019 Aug 21;21(8):1691-1698 [FREE Full text] [doi: 10.1038/s41436-018-0402-0] [Medline: 30573799]
4. de Cock C, Milne-Ives M, van Velthoven MH, Alturkistani A, Lam C, Meinert E. Effectiveness of conversational agents (Virtual assistants) in health care: protocol for a systematic review. JMIR Res Protoc 2020 Mar 9;9(3):e16934. [doi: 10.2196/16934]
5. Perrin A. 10 facts about smartphones as the iPhone turns 10. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/ [accessed 2020-09-02]
6. Attard CA, Carmany EP, Trepanier AM. Genetic counselor workflow study: the times are they a-changin'? J Genet Couns 2019 Feb 10;28(1):130-140. [doi: 10.1002/jgc4.1041] [Medline: 30629774]
7. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M, Working Groups of the National Coordinating Center for Seven Regional Genetics Service Collaboratives. Current conditions in medical genetics practice. Genet Med 2019 Aug 28;21(8):1874-1877 [FREE Full text] [doi: 10.1038/s41436-018-0417-6] [Medline: 30686822]
8. 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]
9. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
10. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
11. 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]
12. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adher 2014 Jan:63. [doi: 10.2147/ppa.s53504]
13. Harper R, Nicholl P, McTear M, Wallace J, Black L, Kearney P. Automated phone capture of diabetes patients readings with consultant monitoring via the web. In: 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS 2008). 2008 Presented at: 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS 2008); March 31 - April 4, 2008; Belfast, UK p. 219-226. [doi: 10.1109/ecbs.2008.31]
14. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30 [FREE Full text] [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
15. Bibault J, Chaix B, Nectoux P, Pienkowsky A, Guillemasse A, Brouard B. Healthcare ex Machina: are conversational agents ready for prime time in oncology? Clin Transl Radiat Oncol 2019 May;16:55-59 [FREE Full text] [doi: 10.1016/j.ctro.2019.04.002] [Medline: 31008379]

16. Owens OL, Felder T, Tavakoli AS, Revels AA, Friedman DB, Hughes-Halbert C, et al. Evaluation of a computer-based decision aid for promoting informed prostate cancer screening decisions among African American men: iDecide. Am J Health Promot 2019 Feb 11;33(2):267-278. [doi: 10.1177/0890117118786866] [Medline: 29996666]
17. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 Aug 21;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
18. Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. J Genet Couns 2019 Dec 24;28(6):1166-1177. [doi: 10.1002/jgc4.1169] [Medline: 31549758]
19. Roberts MC, Kennedy AE, Chambers DA, Khoury MJ. The current state of implementation science in genomic medicine: opportunities for improvement. Genet Med 2017 Aug 12;19(8):858-863 [FREE Full text] [doi: 10.1038/gim.2016.210] [Medline: 28079898]
20. Brito JP, Asi N, Bancos I, Gionfriddo MR, Zeballos-Palacios CL, Leppin AL, et al. Testing for germline mutations in sporadic pheochromocytoma/paraganglioma: a systematic review. Clin Endocrinol (Oxf) 2015 Mar 07;82(3):338-345. [doi: 10.1111/cen.12530] [Medline: 24954084]
21. Currás-Freixes M, Inglada-Pérez L, Mancikova V, Montero-Conde C, Letón R, Comino-Méndez I, et al. Recommendations for somatic and germline genetic testing of single pheochromocytoma and paraganglioma based on findings from a series of 329 patients. J Med Genet 2015 Oct;52(10):647-656. [doi: 10.1136/jmedgenet-2015-103218] [Medline: 26269449]
22. Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, US Multi-Society Task Force on Colorectal Cancer. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on colorectal cancer. Gastroenterology 2014 Aug;147(2):502-526. [doi: 10.1053/j.gastro.2014.04.001] [Medline: 25043945]
23. Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, et al. Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer 2015 Apr 02;15(1):215 [FREE Full text] [doi: 10.1186/s12885-015-1224-y] [Medline: 25886519]
24. Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, et al. Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. J Am Med Assoc 2017 Sep 05;318(9):825-835 [FREE Full text] [doi: 10.1001/jama.2017.11137] [Medline: 28873162]
25. Petrucelli N, Daly M, Pal T. BRCA1-and BRCA2-associated hereditary breast and ovarian cancer. In: GeneReviews®. Seattle: University of Washington; 2016.
26. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol 2017 Oct 20;35(30):3382-3390. [doi: 10.1200/jco.2017.72.3502]
27. Biesecker B. Genetic counseling and the central tenets of practice. Cold Spring Harb Perspect Med 2020 Mar 02;10(3):a038968. [doi: 10.1101/cshperspect.a038968] [Medline: 31570379]
28. Flannery DB. Challenges and opportunities for effective delivery of clinical genetic services in the U.S. healthcare system. Curr Opin Pediatr 2018;30(6):740-745. [doi: 10.1097/mop.0000000000000693]
29. Gordon E, Babu D, Laney D. The future is now: technology's impact on the practice of genetic counseling. Am J Med Genet C Semin Med Genet 2018 Mar;178(1):15-23. [doi: 10.1002/ajmg.c.31599] [Medline: 29512937]
30. Rashkin MD, Bowes J, Dunaway K, Dhaliwal J, Loomis E, Riffle S, et al. Genetic counseling, 2030: an on-demand service tailored to the needs of a price conscious, genetically literate, and busy world. J Genet Couns 2019 Apr 09;28(2):456-465. [doi: 10.1002/jgc4.1123] [Medline: 30964579]
31. Portela M, Granell-Canut C. A new friend in our smartphone? Observing interactions with chatbots in the search of emotional engagement. In: Proceedings of the XVIII International Conference on Human Computer Interaction. 2017 Presented at: Interacción '17: XVIII International Conference on Human Computer Interaction; September 25 - 27, 2017; Cancun Mexico p. 1-7. [doi: 10.1145/3123818.3123826]
32. Chaves AP, Gerosa MA. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. Int J Hum Comput Interact 2020 Nov 08;37(8):729-758. [doi: 10.1080/10447318.2020.1841438]
33. Genetic/familal high risk assessment: breast and ovarian. National Comprehensive Cancer Network. 2019. URL: https://www2.tri-kobe.org/nccn/guideline/gynecological/english/genetic_familial.pdf [accessed 2020-11-11]
34. Genetic/familal high risk assessment: colorectal. National Comprehensive Cancer Network. 2020. URL: https://jnccn.org/configurable/content/journals$002fjnccn$002f17$002f9$002farticle-p1032.xml?t:ac=journals%24002fjnccn%24002f17%24002f9%24002farticle-p1032.xml [accessed 2020-11-11]
35. Del Fiol G, Kohlmann W, Bradshaw RL, Weir CR, Flynn M, Hess R, et al. Standards-based clinical decision support platform to manage patients who meet guideline-based criteria for genetic evaluation of familial cancer. JCO Clin Cancer Informat 2020 Sep(4):1-9. [doi: 10.1200/cci.19.00120]
36. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge University Press; 2014:43-71.
37. Knowles M. Andragogy, not pedagogy. Adult Leadership 1968;16(10):350-352 [FREE Full text]

38. Tegos S, Demetriadis S. Conversational agents improve peer learning through building on prior knowledge. J Edu Technol Soc 2017;20(1):99-111 [FREE Full text]
39. Wang C, Bickmore T, Bowen DJ, Norkunas T, Campion M, Cabral H, et al. Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories. Genet Med 2015 Oct 15;17(10):822-830 [FREE Full text] [doi: 10.1038/gim.2014.198] [Medline: 25590980]
40. Pyeritz RE. The family history: the first genetic test, and still useful after all those years? Genet Med 2011;14(1):3-9. [doi: 10.1097/gim.0b013e3182310bcf]
41. Heald B, Edelman E, Eng C. Prospective comparison of family medical history with personal genome screening for risk assessment of common cancers. Eur J Hum Genet 2012 May 4;20(5):547-551 [FREE Full text] [doi: 10.1038/ejhg.2011.224] [Medline: 22215421]
42. Guttmacher AE, Collins FS, Carmona RH. The family history — more important than ever. N Engl J Med 2004 Nov 25;351(22):2333-2336. [doi: 10.1056/nejmsb042979]
43. Yoon PW, Scheuner MT, Peterson-Oehlke KL, Gwinn M, Faucett A, Khoury MJ. Can family history be used as a tool for public health and preventive medicine? Genet Med 2002 Aug;4(4):304-310. [doi: 10.1097/00125817-200207000-00009] [Medline: 12172397]
44. Wilson BJ. Systematic review: family history in risk assessment for common diseases. Ann Intern Med 2009 Dec 15;151(12):878-885. [doi: 10.7326/0000605-200912150-00177]
45. Giovanni MA, Murray MF. The application of computer-based tools in obtaining the genetic family history. Curr Protoc Hum Genet 2010 Jul;66(1):9-21. [doi: 10.1002/0471142905.hg0921s66] [Medline: 20582917]
46. My family health portrait - public health genomics knowledge base. Centers for Disease Control and Prevention. URL: https://phgkb.cdc.gov/FHH/html/index.html [accessed 2021-03-22]
47. Bevers TB, Helvie M, Bonaccio E, Calhoun KE, Daly MB, Farrar WB, et al. Breast cancer screening and diagnosis, version 3.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2018 Nov 15;16(11):1362-1389. [doi: 10.6004/jnccn.2018.0083] [Medline: 30442736]
48. Provenzale D, Gupta S, Ahnen DJ, Markowitz AJ, Chung DC, Mayer RJ, et al. NCCN guidelines insights: colorectal cancer screening, version 1.2018. J Natl Compr Canc Netw 2018 Aug 11;16(8):939-949. [doi: 10.6004/jnccn.2018.0067] [Medline: 30099370]
49. Williams CD, Grady WM, Zullig LL. Use of NCCN guidelines, other guidelines, and biomarkers for colorectal cancer screening. J Natl Compr Canc Netw 2016 Nov 31;14(11):1479-1485 [FREE Full text] [doi: 10.6004/jnccn.2016.0154] [Medline: 27799515]

Abbreviations

AI: artificial intelligence
EHR: electronic health record

Edited by R Kukafka; submitted 07.04.21; peer-reviewed by D Cragun, K Kalke; comments to author 28.06.21; revised version received 22.08.21; accepted 12.09.21; published 18.11.21.

Please cite as:
Chavez-Yenter D, Kimball KE, Kohlmann W, Lorenz Chambers R, Bradshaw RL, Espinel WF, Flynn M, Gammon A, Goldberg E, Hagerty KJ, Hess R, Kessler C, Monahan R, Temares D, Tobik K, Mann DM, Kawamoto K, Del Fiol G, Buys SS, Ginsburg O, Kaphingst KA
Patient Interactions With an Automated Conversational Agent Delivering Pretest Genetics Education: Descriptive Study
J Med Internet Res 2021;23(11):e29447
URL: https://www.jmir.org/2021/11/e29447
doi:10.2196/29447
PMID:34792472

Original Paper

Effects of Emotional Expressiveness of a Female Digital Human on Loneliness, Stress, Perceived Support, and Closeness Across Genders: Randomized Controlled Trial

Kate Loveys[1], MSc; Mark Sagar[2,3], PhD; Xueyuan Zhang[2], PhD; Gregory Fricchione[4,5], MD; Elizabeth Broadbent[1], PhD

[1]Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand

[2]Soul Machines Ltd, Auckland, New Zealand

[3]Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand

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

[5]Benson-Henry Institute of Mind Body Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Elizabeth Broadbent, PhD
Department of Psychological Medicine
The University of Auckland
Building 507, Level 3
22-30 Park Avenue
Auckland, 1023
New Zealand
Phone: 64 9 923 6756
Email: e.broadbent@auckland.ac.nz

Abstract

Background: Loneliness is a growing public health problem that has been exacerbated in vulnerable groups during the COVID-19 pandemic. Social support interventions have been shown to reduce loneliness, including when delivered through technology. Digital humans are a new type of computer agent that show promise as supportive peers in health care. For digital humans to be effective and engaging support persons, it is important that they develop closeness with people. Closeness can be increased by emotional expressiveness, particularly in female relationships. However, it is unknown whether emotional expressiveness improves relationships with digital humans and affects physiological responses.

Objective: The aim of this study is to investigate whether emotional expression by a digital human can affect psychological and physiological outcomes and whether the effects are moderated by the user's gender.

Methods: A community sample of 198 adults (101 women, 95 men, and 2 gender-diverse individuals) was block-randomized by gender to complete a 15-minute self-disclosure conversation with a female digital human in 1 of 6 conditions. In these conditions, the digital human varied in modality richness and emotional expression on the face and in the voice (emotional, neutral, or no face; emotional or neutral voice). Perceived loneliness, closeness, social support, caring perceptions, and stress were measured after each interaction. Heart rate, skin temperature, and electrodermal activity were assessed during each interaction. 3-way factorial analyses of variance with post hoc tests were conducted.

Results: Emotional expression in the voice was associated with greater perceptions of caring and physiological arousal during the interaction, and unexpectedly, with lower feelings of support. User gender moderated the effect of emotional expressiveness on several outcomes. For women, an emotional voice was associated with increased closeness, social support, and caring perceptions, whereas for men, a neutral voice increased these outcomes. For women, interacting with a neutral face was associated with lower loneliness and subjective stress compared with no face. Interacting with no face (ie, a voice-only black screen) resulted in lower loneliness and subjective stress for men, compared with a neutral or emotional face. No significant results were found for heart rate or skin temperature. However, average electrodermal activity was significantly higher for men while interacting with an emotional voice.

Conclusions: Emotional expressiveness in a female digital human has different effects on loneliness, social, and physiological outcomes for men and women. The results inform the design of digital human support persons and have theoretical implications.

Further research is needed to evaluate how more pronounced emotional facial expressions in a digital human might affect the results.

Trial Registration: Australia New Zealand Clinical Trials Registry (ANZCTR) ACTRN12621000865819; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=381816&isReview

(J Med Internet Res 2021;23(11):e30624) doi:10.2196/30624

KEYWORDS

computer agent; digital human; emotional expressiveness; loneliness; closeness; social support; stress; human-computer interaction; voice; face; physiology

Introduction

Background

Loneliness is a growing public health issue that has been exacerbated in vulnerable groups during the COVID-19 pandemic [1]. It is a subjective psychological state in which a person perceives a discrepancy between the quality and quantity of their actual and ideal social relations [2]. It has been described as a *syndrome-like* condition that reliably affects social cognition and behavior in a manner that perpetuates feelings of loneliness [3-5]. Loneliness has been shown to be most prevalent in young adults (aged 18-29 years); older adults; people with a physical or mental health condition; people with low income; people who live alone; and people who are separated, widowed, or divorced [1,6].

Loneliness is a particularly pressing public health problem because it has been associated with a multitude of ill health effects, which places a burden on health care systems [3,7]. Loneliness produces a greater mortality risk than smoking 15 cigarettes every day [8], and it increases the risk of many physical and mental morbidities. This includes a greater risk of coronary heart disease, stroke [9], and psychiatric conditions, including major depressive disorder, generalized anxiety disorder, and suicide [10,11]. Interventions are needed to alleviate the burden of loneliness on population health and health care systems.

Social Support Interventions

Social support has been shown to be a major protective factor against loneliness in adults [6,12] and an effective intervention strategy for loneliness [4]. In a survey of 38,217 adults in the United Kingdom, a high level of social support was associated with 89% lower odds of developing severe loneliness [12]. Moreover, reporting ≥3 close relationships was associated with 42% lower odds of severe loneliness. This suggests that interventions for improving social support and close relationships may help combat loneliness, which has been found in a systematic review of loneliness interventions [4].

Social support is a construct that refers to the structure or function of interpersonal relationships [13]. Structural social support describes aspects of a person's social network, including the existence of relationships (eg, number of friends) or the types of connections between people (eg, spousal). Functional social support describes the functions of a particular relationship or interaction for a person (eg, emotional, informational, instrumental, or appraisal support). Both structural and functional support have been associated with reductions in loneliness to varying degrees [5]. However, satisfaction with functional support that has been received has been found to have the strongest relationship with loneliness. This suggests that social support interventions focused on improving functional support could have the strongest effect on loneliness.

Psychological interventions have been shown to increase social support [14]. This may be through one of several techniques including teaching skills to help people strengthen their support networks or to better solicit functional support, or the direct provision of functional support to people (eg, emotional support from a support group). Psychological interventions for improving social support have been shown to reduce loneliness [4] and improve health outcomes [14]. Social support may improve health outcomes indirectly through the provision of functional support (eg, transportation to a health appointment) or directly by exerting a stress-buffering effect [15,16]. Social support has been associated with reduced sympathetic nervous system activation [17] and reduced cortisol and increased oxytocin release [18], which may help to buffer against the effects of loneliness-related stress on the body. Chronic stress has been associated with reduced heart rate variability [19], increased inflammation [17,20], and impaired immune response [21], which increases the risk of physical and mental health conditions.

Social support may have stress-buffering and loneliness-reducing effects because humans have evolved to survive in groups. Owing to this evolution, humans experience separation distress and find relief in their social attachments [22]. Social support is an *attachment solution* that is encouraged in mammalian brain structure and neural processing loops (eg, the *protolimbic* and *paralimbic cortico-striato-thalamocortical* loops) to provide solace and alleviate distress. Attachment solutions may be found in other people, animals, religious figures, and transitional objects (eg, a child's teddy bear) [22]. In-person social support interventions may not always be available (eg, because of physical restrictions during a pandemic, hospitalization, and living rurally). When faced with the absence of human connection, artificial agents (such as robots or computer agents) might provide digital social support to people that may be salutogenic in the context of loneliness.

Robot and Computer Agent Support

Research has increasingly shown that social robots and computer agents may help to reduce loneliness and improve health outcomes for people by providing functional social support. A social robot is a technology with a body in physical hardware

form, which is programmed to perceive and act autonomously within its physical environment and is capable of social interaction [23]. For example, Paro is a fluffy baby harp seal robot that includes artificial intelligence for social interactions and a heater for body warmth. Paro may provide emotional support to people through companionship and physical touch akin to a pet. Paro has been shown to reduce loneliness in older adults, and reduce pulse rate and blood pressure [24-26]. Other robots have reduced loneliness by facilitating interactions between people and their support networks over video calls [27,28].

A computer agent refers to a computer-generated entity that may include a dialog system and an embodiment (eg, an animation of a face or body) [29,30]. Computer agents are often capable of social interaction and may include technologies such as embodied conversational agents, chatbots, game characters, and digital humans. There is limited research investigating the effect of support from a computer agent on loneliness [31]. However, 1 study found that companionship from an animal-like conversational agent reduced loneliness in hospitalized older adults [32]. Other studies have found that Tanya, a humanlike computer agent companion for older adults at home, was effective at improving loneliness and was highly acceptable to users [33,34]. Similarly, studies in young adult populations have found that computer agents show promise for improving loneliness [35,36].

Digital humans are a new form of computer agent that show promise for applications in health care, including acting as a supportive peer. Digital humans are computer-based, autonomous animations in the form of a human face or a full body. They use complex neurobehavioral modeling techniques involving virtual neurotransmitters and a visual computing framework described by Sagar et al [37]. These techniques enable emotional intelligence, personality, and complex social interactions. For example, digital humans can show attachment and separation distress toward a user while in *high oxytocin mode*. *High oxytocin mode* is a setting that involves more rapid firing of virtual oxytocin in the digital human's autonomous brain model when a human is detected in front of the computer's webcam [38]. This setting influences the digital human's emotional expressions to become increasingly distressed when the user goes out of view and become more positive when the user returns. As digital humans exist on a screen, they may be better suited to providing emotional and informational support to people as opposed to instrumental support (eg, helping with physical tasks).

Digital humans (and computer agents in general) are a new technology, and research is needed to understand how to design them in a way that is conducive to providing support to people. Research has shown that the perceived effectiveness of social support interventions may be influenced by how socially close people feel to the support partner [39]. This suggests that it may be important to design digital humans to engage in behaviors that build closeness with people while providing support. However, there is limited research on which behaviors increase closeness with computer agents [40].

Improving Relationships With Computer Agents

Psychological research has suggested that closeness can be developed in human relationships using several techniques. These include engaging in reciprocal self-disclosure, undertaking shared activities or interests, expressing the value of the relationship, and showing empathy, among others [41,42]. However, there may be gender differences in the techniques that are most important for building closeness [41,43,44]. Women may place greater importance on emotional self-disclosure in close relationships, whereas men may find shared ideas and hobbies to be more important to closeness [41,44]. It is unclear whether behavioral techniques that increase closeness in human relationships translate to human and computer agent relationships; however, the computers are social actors (CASA) paradigm would suggest so [45].

CASA posits that people interact with computers that provide social cues in a manner similar to how they would with another person [46]. Findings from several experiments support CASA [45], showing that people make personality judgments [47] and gender stereotypes [48], engage in mindless social behaviors such as reciprocity [45], and elicit in-group and outgroup behaviors toward computers [49]. More recently, studies have shown that CASA applies to advanced technologies. A voice agent in an autonomous vehicle was perceived more positively when its communication style conformed to gender stereotypes (eg, a sociable female voice vs an informative male voice) [50]. Another study found that the similarity-attraction principle applied to a voice agent in a smart-home environment [51]. Extroverted users were found to prefer a talkative voice agent, whereas introverted users preferred an agent with multiple voices because it felt like talking to several less-talkative agents. According to CASA, computer agents that display closeness-building behaviors should form closer relationships with people than agents that do not.

Emotional Expressiveness

Emotional expressiveness is a technique that can increase closeness in human relationships, particularly in female relationships [41,44]. Emotional expressiveness refers to expressing an intimate degree of negative or positive affect (eg, through facial expression, speech, and gaze), usually while engaging in behaviors such as talking at a personal level about fears or personal problems, or while demonstrating care [44]. Emotional expressiveness may help build emotional intimacy, a type of psychological intimacy that is important in close relationships [52]. It may also be a way of providing nurturance, which is one of the 3 types of social attachment–building behaviors in mammals, according to the Mammalian Behavioural Triad theory [53]. As digital humans have a detailed animated face modeled on human musculature and a voice, they are capable of engaging in multimodal emotional expression. This involves delivering congruent emotional cues through different communication modalities, such as facial expressions and speech tone.

Although promising, research has yet to investigate whether emotional expressiveness in a digital human increases feelings of closeness toward it. Related research on social robots has found gender differences in the effect of emotional

expressiveness on closeness. Women have been shown to feel closer to a social robot that expresses emotions on the face during interactions (alongside other relational behaviors, eg, self-disclosure, mirrored posture, and speaking rate), whereas men reported feeling closer to robots that expressed fewer emotions and relational behaviors [54]. It is possible that similar effects could be found with computer agents; however, experimental research is needed to investigate this.

Prior research on computer agents has shown that emotional expressiveness can increase social perceptions related to closeness. A computer agent that expressed empathy on the face and in the voice was rated as significantly more supportive, caring, and likable compared with an agent that displayed more self-focused emotion (eg, joy at its own wins in a game) [55]. Similarly, another computer agent that displayed a range of emotions on the face and in the voice was perceived as significantly more warm, caring, cooperative, and trustworthy than an agent with no emotion [56].

These studies show that emotional expressiveness in a computer agent may increase perceptions related to closeness; however, no studies have looked at the effects on closeness directly. In addition, from the existing research, it is unclear whether emotions expressed through the face or voice are more important for closeness building and supportive interactions with a computer agent. Some studies have shown that the presence of a face may help to increase attentional engagement with a computer agent [57], which could improve how well users attend to emotional cues. However, it remains unclear whether the presence of a computer agent face can affect the development of social closeness. Several instruments have been developed to quantify closeness, including the Inclusion of Other in the Self scale [58], the Perceived Interpersonal Closeness Scale [59], and the Relationship Closeness Induction Task (RCIT) measure (for experimentally induced feelings of closeness) [60].

This Study

Research has yet to investigate the effects of emotional expressiveness in a digital human's face and voice on user outcomes, and how emotional expressiveness interacts with user gender during a supportive interaction. The aim of this study is to investigate the effect of emotional expressiveness in a female digital human on loneliness, closeness, caring perceptions, social support, stress, and physiological arousal in a community sample. It was hypothesized that there would be gender differences in the effect of emotional expressiveness in a female digital human on outcomes. Particularly, it was anticipated that women would report the greatest reductions in loneliness, stress, and physiological arousal and the greatest increases in social support, closeness, and caring perceptions in response to a female digital human with an emotional face and an emotional voice. In contrast, it was predicted that men would report better outcomes in response to a female digital human with a neutral face and a neutral voice.

Methods

Study Design

A between-group experimental study was conducted to investigate the effect of multimodal emotional expression and user gender on loneliness after a self-disclosure interaction with a female digital human. Secondary outcomes included social closeness, caring perceptions, social support, stress, and physiological arousal. Participants were block-randomized by gender to one of 6 conditions in which the digital human's design differed by face type (emotional, neutral, or no face) and voice type (emotional or neutral; described in the *Digital Human* section).

The study procedures were approved by the University of Auckland Human Participants Ethics Committee on November 1, 2018 (reference number: 022191). Retrospective registration was provided by the Australia New Zealand Clinical Trials Registry (registration number: ACTRN12621000865819). Retrospective registration was sought as the study was conceived of as a human-computer interaction experiment (to understand the effect of a digital human's design) as opposed to a clinical trial. However, later registration was sought, given the number of psychological and physiological outcomes collected as part of the experiment. This study is reported in keeping with the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines [61].

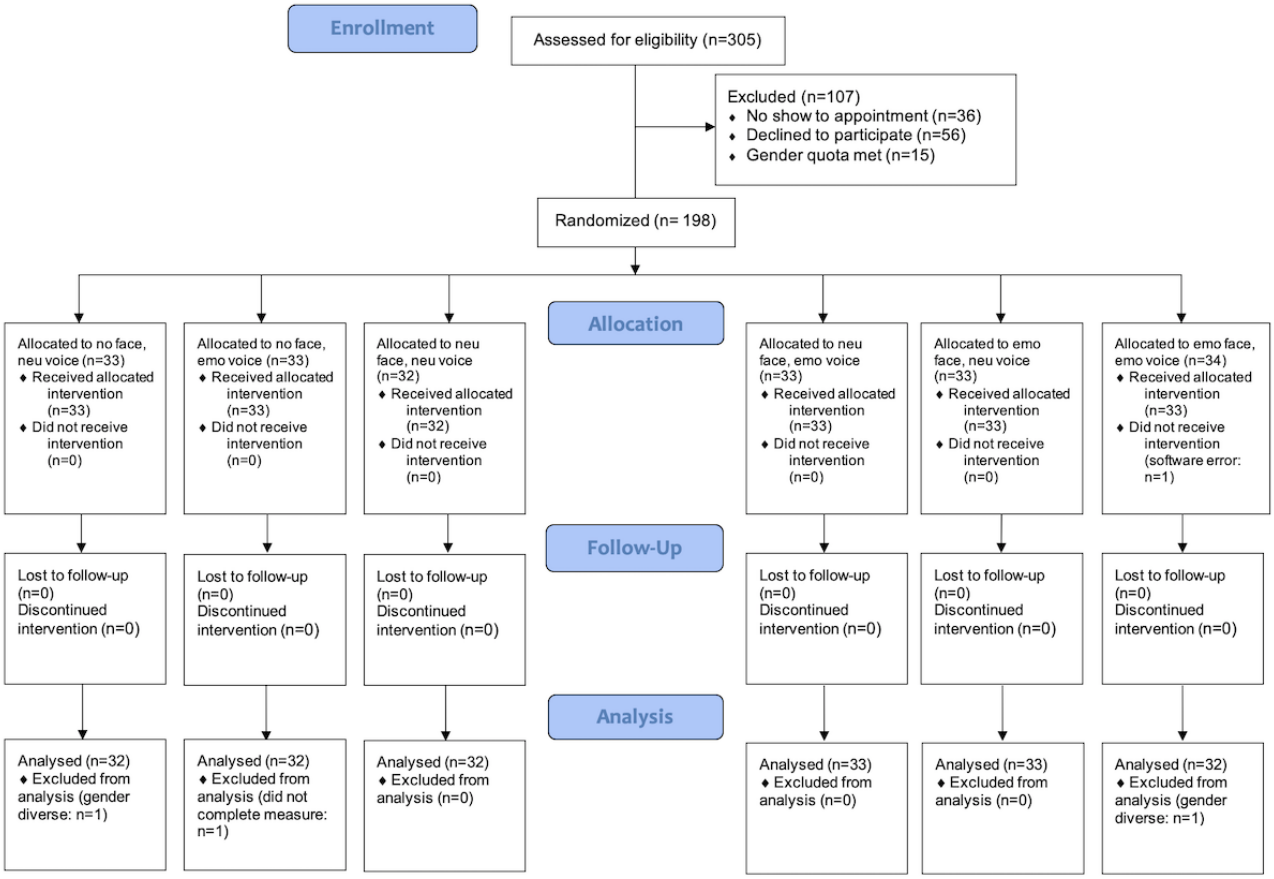
Participant Recruitment

A community sample of 198 adults (101 women, 95 men, and 2 gender-diverse individuals) was recruited to participate in a study on relationships with a digital human. Participants were considered eligible if they were aged ≥18 years and were fluent in English. The study was advertised using flyers distributed in the University of Auckland campuses and through Facebook advertising to the Auckland city area. Compensation of a NZ $20 (US $14.33) shopping voucher was offered for participation. Participants were recruited between February 20, 2019, and July 24, 2019.

A power analysis was conducted using G*Power software to determine the required sample size. This was informed by the results of Brave et al [55], who compared the effects of agent empathy versus no empathy on feelings of support and found an effect size of f=0.32. 198 participants would be needed to detect an effect size of f=0.32, with 80% power and an α level of .05 for a 3 (face) by 2 (voice) by 2 (gender) analysis of variance (ANOVA). A recruitment target was set for 100 women and 100 men.

Data from 1.5% (3/198) of participants were excluded from analyses for the following reasons: the software would not start (1/198, 0.5%) or the participant was identified as gender diverse, and this group was not of sufficient size for adequate statistical power in gender analyses (2/198, 1%). As a result, data from 195 participants were included in the analyses. A CONSORT diagram depicts participant flow through the study (Figure 1).

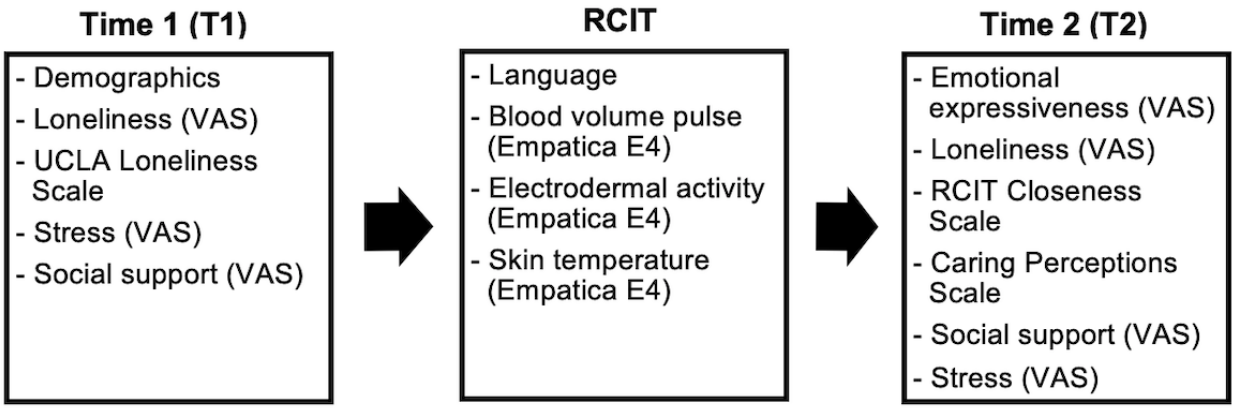
Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram (emo: emotional; neu: neutral).



Measures

Figure 2 depicts the time points at which the measures were administered and data were collected in the study.

Figure 2. Time points for administration of measures and data collection. RCIT: Relationship Closeness Induction Task; VAS: visual analog scale.



Demographics

Demographic variables including age, gender, ethnicity, marital status, and occupation were assessed by self-report at baseline. Gender was assessed using a single-response, multiple-choice question with response options (male, female, gender-diverse, and prefer not to divulge). Other demographic variables were evaluated using an open-ended self-report item (age) and multiple-choice questions with response options (ethnicity, marital status, and occupation).

Perceived Emotional Expressiveness Manipulation Check

A manipulation check evaluated the extent to which the digital human was perceived as emotionally expressive using a visual analogue scale. Participants marked their answers by placing an X at the appropriate position along a 100-mm line with

response anchors at either end (0=*not expressive*; 100=*very expressive*). A score was derived by measuring the distance of the X along the line from the left-hand side. A higher score indicated that the digital human was perceived as more emotionally expressive. Perceived emotional expressiveness was measured immediately after the digital human interaction.

Loneliness

General loneliness was measured at baseline using the 20-item Revised UCLA Loneliness Scale [62]. Participants were asked to rate how often they felt the way described in general across a range of statements using a 4-point Likert scale (1=*never*; 4=*often*). The Revised UCLA Loneliness Scale has been shown to have good internal consistency reliability (Cronbach α=.94), discriminant validity with mood measures, and concurrent validity with measures of social isolation (eg, feelings of isolation, time spent alone per day, and number of close friends) [62]. Responses were summed to derive a total loneliness score. A higher score indicated greater loneliness.

State loneliness was evaluated using a visual analogue scale with response anchors placed at each end of the scale (0=*not at all*; 100=*extremely*). Participants marked how lonely they currently felt by placing an X at the appropriate place along a 100-mm line. Scores were derived by measuring the distance of the X along the line. A higher score indicated that the participant experienced a greater degree of loneliness. A visual analogue scale was chosen as it would be more sensitive to changes in feelings of loneliness within a 1-hour experiment session compared with a general loneliness measure.

Social Closeness

Social closeness with the digital human was assessed using the RCIT follow-up measure [60]. This 4-item scale evaluates the degree of perceived relationship closeness, similarity, liking, and likelihood of a future friendship using a 9-point scale. Responses range from 0 (*not at all*) to 9 (*very*). Scores were derived by measuring the distance of an X placed along a 100 mm line and transforming this value to a score from 0 to 9. Scores were summed for the 4 items. A higher score indicated stronger social closeness to the conversation partner. The scale demonstrated good internal consistency reliability in this study (Cronbach α=.86).

Perceived Social Support

Perceived social support was measured using a visual analogue scale with response anchors (0=*not at all*; 100=*extremely*) [63]. Participants rated their extent of agreement with the statement "I feel supported" by placing an X at the appropriate spot along a 100-mm line. The distance of the X along the line was measured, and a score was derived. Higher scores indicated feeling more supported. A visual analogue scale was chosen as it may be more sensitive to changes in perceived support after a short experiment session than a general measure. Although lower in internal reliability compared with a multidimensional scale, the 1-item visual analogue scale helped to reduce participant burden and was thought to be suitable for assessing feelings of support after a brief interaction.

Caring Perceptions

Caring perceptions are judgments about an agent's traits (whereas feeling supported pertains to the effect of the interaction on the user). Caring perceptions were assessed using the Caring Perceptions Scale by Brave et al [55]. This is a 5-item scale that evaluates the extent to which a computer agent is perceived as caring. Items are rated on a 10-point scale with semantic differential response anchors (eg, *not compassionate* to *compassionate*). Items in the scale evaluated how warm, compassionate, selfish, friendly, and cooperative the digital human was. This scale has been used in prior research looking at relationships with embodied conversational agents and has shown good internal consistency reliability in university student samples (Cronbach α=.88) [55]. Items were summed, and a higher score indicated that the participant perceived the digital human as more caring.

Perceived Stress

Current stress was measured using a 100-mm visual analogue scale with response anchors (1=*not at all*; 100=*extremely*) [64,65]. Participants rated their extent of agreement with the phrase *"I feel stressed"* by placing an X at the appropriate point along the line. The distance from the left edge of the line was measured, and a score was derived. A higher score indicated greater perceived stress. This visual analogue scale was chosen because it was shown to be sensitive to changes in perceived stress after a brief experimental session in previous research [64,65].

Physiological Stress Response

Heart rate, electrodermal activity, and skin temperature are measures of physiological arousal that can indicate stress [66,67]. Heart rate, electrodermal activity, and skin temperature were measured using a wrist-worn sensor device called Empatica E4 (Empatica Inc). Data were continuously collected during the interaction and subsequently processed using a Python script to create average heart rate (converted from blood volume pulse data), electrodermal activity, and skin temperature scores per participant. Higher average scores in heart rate and electrodermal activity and lower scores in wrist skin temperature indicated greater physiological arousal during the interaction with the digital human. Data were analyzed from the beginning to the end of the participant's interaction with the digital human. Data from only 170 participants were included in the analyses for physiological stress, as 25 data files were lost to processing errors.

Procedure

Participants attended a 45-minute experimental session at the University of Auckland Clinical Research Centre. After providing written informed consent, participants secured an Empatica E4 sensor to their wrist and completed a baseline questionnaire on paper. The baseline questionnaire assessed demographic and psychological variables, including loneliness, stress, and social support. Participants were then block-randomized by gender to interact with 1 of 6 versions of a digital human. Block randomization was completed before the session by a member of the research team, who automatically generated a randomization table using Research Randomizer

software. Allocations were concealed from the researcher and the participant in opaque envelopes. The researcher remained blinded to the participant's condition until the envelope was opened immediately before starting the appropriate computer program for the participant. Although the participants were deblinded to their condition upon starting their interaction, they remained unaware of what digital humans in the other experimental conditions were like. The 6 digital humans varied in terms of modality richness and emotional expression (emotional face, neutral face, no face; emotional voice, and no voice).

The researcher provided the participant with verbal and written instructions for interacting with the digital human, started the digital human program on the laptop computer, then closed the door, and exited the room for the duration of the interaction. Participants were instructed to call on the researcher (who was sitting on a chair down the hallway) by ringing a loud desk bell if they encountered any technical difficulties that they could not resolve.

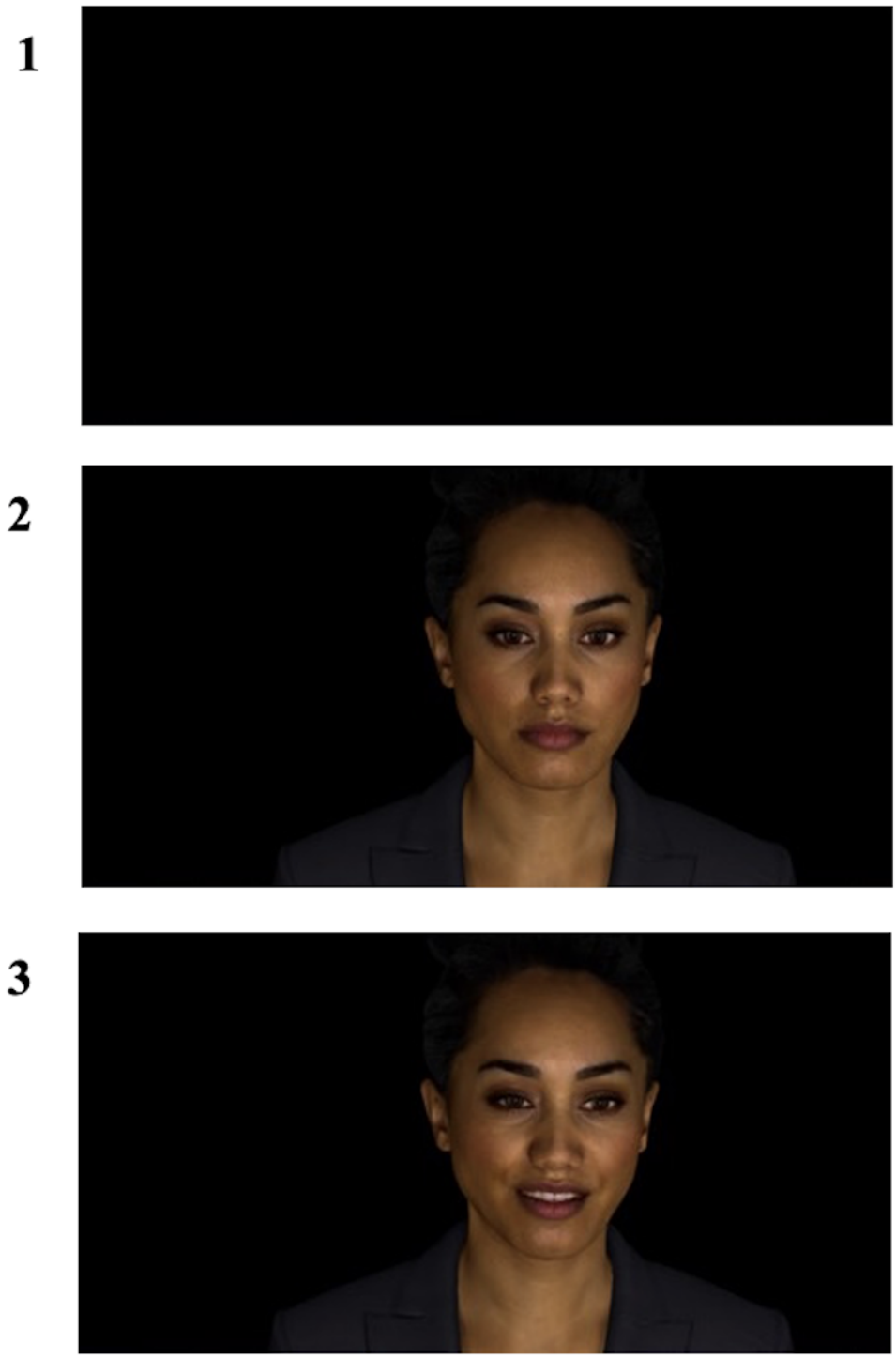
Participants completed the RCIT with a digital human. The RCIT is a 15-minute structured conversation involving reciprocal self-disclosure of personal information, feelings, and memories over a list of 28 personal questions [60]. Questions gradually increase in intimacy and cover topics from "What is your name?" to "Describe the last time you felt lonely." The RCIT has been shown to reliably induce a sense of closeness among human-stranger dyads in experimental psychology research [60], and it has been shown to promote health benefits found in naturally occurring close relationships, including faster wound healing speed and reduced stress [64]. The RCIT is divided into 3 sections with a time limit for each (2, 5, and 8 minutes). To maintain a private interaction between the participant and the digital human, the participant was required to ring a loud desk bell each time they reached the end of a section. The researcher sat in the hallway running a digital timer that was cleared at the beginning of each section. If a participant went over the time limit for a given section, the researcher knocked on the door and the participant was instructed to move on to the next section. Participants' language from their interaction with the digital human was recorded and analyzed for emotional content in another paper [68]. Once the RCIT was completed, the participant rang the desk bell, and the researcher returned to the room. The participant removed the wrist sensor and completed a second paper questionnaire on perceived loneliness, closeness, caring perceptions, social support, and stress. Participants also answered a series of open-ended, written qualitative questions on their perceptions of the digital human (reported in another paper; Loveys, unpublished data, March 2021). Once the questionnaire was completed, the participant was provided with a NZ $20 (US $14.33) shopping voucher.

Digital Human

Overview

The digital human was presented on a laptop computer screen in a portrait view with an embodiment of the head and shoulders depicted in front of a black background. The digital human was named Holly, and she was modeled on a young adult female of mixed ethnicity (New Zealand Māori, New Zealand European, Indian, and European). In the *no-face* conditions, the digital human appeared as a black screen only. Figure 3 illustrates how the digital human appeared in the no-face, neutral face, and emotional face conditions. The digital human functioned autonomously using a finite-state conversation engine and a neurobehavioral modeling and visual computing framework, as described by Sagar et al [37].

Figure 3. The digital human across the 3 visual conditions (1: no face; 2: neutral face; 3: emotional face). Each visual condition could be paired with either an emotional or neutral voice.



Before programing the digital human's emotional expressions, a member of the research team (KL) recruited 4 psychology master's students to complete the RCIT in 2 dyads (1 female-female; 1 male-male). Each dyad completed the task separately in a private room with the researcher present. The researcher recorded notes on the types of emotions expressed by the dyads during each of the 28 questions in the conversation task. Notes were also made regarding the types of responses people gave to different questions. This information was used to inform the answers the digital human gave to the RCIT questions and the types of emotions it expressed on the face and in the voice for each question. This approach was taken to improve how natural the digital human's responses were. The digital human's language was consistent between participants, as were the emotional expressions on the face and in the voice for individuals in those respective conditions. This design choice

was made to maintain experimental control against the potential confounds of different conversation content or amounts of emotional expression, which could have affected the outcomes.

The digital human's script underwent an iterative development process involving the research team drafting responses to the RCIT questions, programming responses into the digital human as part of a finite-state conversation engine, and then testing the conversation with other members of the research team. The digital human's responses were refined over several iterations to make them seem natural and convincing. The digital human interacted autonomously with participants.

Emotional Face

The emotional face condition included a mixture of positive and negative expressions, including compassion, joy, fear, and sadness. Compassion and joy were expressed most frequently (113/213, 53% and 82/213, 38.5% of the interactions, respectively). The emotional expressions were subtle in order to seem more natural. The emotional expressions on the face were preprogrammed and triggered based on the phrases that the digital human spoke using a text-to-speech Emotional Markup Language. Programing the Emotional Markup Language required manually classifying each sentence the digital human spoke into compassion, joy, fear, or sadness categories in the digital human program. The result was that each time the digital human spoke a phrase, she elicited the associated emotional expression on her face. Emotional expressions on the face were autonomously generated in real time as the digital human spoke to the participant using neurobehavioral modeling and visual computing techniques. These techniques are described in detail by Sagar et al [37,38]. While listening, the digital human did not express emotions on the face, although she engaged in humanlike facial movements, such as blinking, as described in the *Other Behaviors* section.

The neutral face condition contained the same digital human face as the emotional face condition; however, no facial expressions were made during the interaction. The only movements the neutral face made were of the mouth to accommodate speech and the natural behaviors described in the *Other Behaviors* section (eg, blinking and head movements with speech). In the no-face condition, participants interacted with a plain black screen that contained no digital human animation and only a voice.

The digital human was capable of mirroring people's facial expressions; however, this capability was turned off for the experiment. This was to ensure that the digital human did not provide emotional facial expressions in the neutral face condition and that participants in the emotional face condition received the same amount of emotional expression from the digital human. However, the tradeoff was that the digital human could not return a person's smile, for example.

Emotional Voice

Participants interacted with a digital human that had either an emotional or neutral voice. The digital human's voice was recorded by a young adult female voiceover artist with a local New Zealand accent. Emotions expressed in the emotional voice condition were compassion (ie, low arousal positive affect; 113/213 statements, 53% of the conversation), joy (ie, high arousal positive affect; 82/213 statements, 38.5% of the conversation), and sadness or anxiety (ie, low arousal negative affect; 18/213 statements, 8.5% of the conversation). The neutral voice had an absence of positive and negative emotions; however, it maintained normal speech intonation for a New Zealand accent. Each sentence in the digital human's script was recorded twice—one in a neutral voice and the other in an emotional voice. For the emotional voice, the voiceover artist read from a script with annotations to guide when certain emotional expressions should take place. These emotional expressions were tied to phrases to correspond with when emotions were portrayed in the digital human's face. Recording the voice clips was an iterative process where a member of the research team (KL) and an audio engineer provided subjective feedback to the vocal artist at the time of recording, and the vocal artist rerecorded clips where necessary. Clips were rerecorded when the emotional expression was either too strong or not strong enough or where there appeared to be emotion in a neutral voice clip.

To ensure that the emotional and neutral voices were sufficiently different, an emotion classifier was used to objectively analyze the vocal clips. The classifier operated with 79.9% accuracy and classified vocal clips based on energy and pitch contours, which were modeled using Gaussian mixture models. Analyses revealed that the compassionate and joyful emotional voice clips were significantly different from the neutral voice clips. However, there were insufficient sadness and anxiety voice clips to reach significance. The voice recordings were generated in a voice package that was connected to a finite-state conversation engine.

Other Behaviors

The digital human engaged in several other behaviors to appear more humanlike during the interaction. She maintained eye contact with people for the majority of the interaction using computer vision technology. Before speaking, she would often look at the top-right of the screen as if thinking. She also engaged in humanlike movements, such as blinking, moving her head, raising her eyebrows with speech, and lightly swaying her body while listening or speaking. If she did not understand the participant, she would say, "I didn't understand, you can try rephrasing." The digital human did not perform any other behaviors during the interaction.

Data Analysis

Data were analyzed using SPSS software, version 27 (IBM Corporation). Data were checked for violations of test assumptions, and bootstrapping was applied to tests where data were not normally distributed. Chi-square tests and ANOVA tests were conducted to check for baseline group differences in demographic and psychological variables. As no significant differences were found in the baseline variables, they were not controlled for in subsequent analyses. A series of 3-way factorial ANOVA tests were conducted to evaluate the effect of face type, voice type, and gender on outcomes at time point 2 (T2). Post hoc tests with Sidak correction were applied as follow-up analyses.

Results

Participants

Participants were predominantly young adults (mean 28.31, SD 10.97), single (96/195, 49.2%), and university students (132/195, 67.7%). There was an approximately equal number of men (94/195, 48.2%) and women (101/195, 51.8%). Participants represented a mix of ethnicities (80/195, 41.0% Asian; 67/195, 34.4% New Zealand European; 38/195, 19.5% other; 6/195, 3.1% Māori; and 4/195, 2.1% Pacific Peoples). Participants reported a moderate degree of loneliness at baseline (mean 36.97, SD 10.04). There were no significant differences in baseline demographic or psychological variables between the experimental groups, and there were no significant differences in psychological variables between men and women at baseline.

Manipulation Check

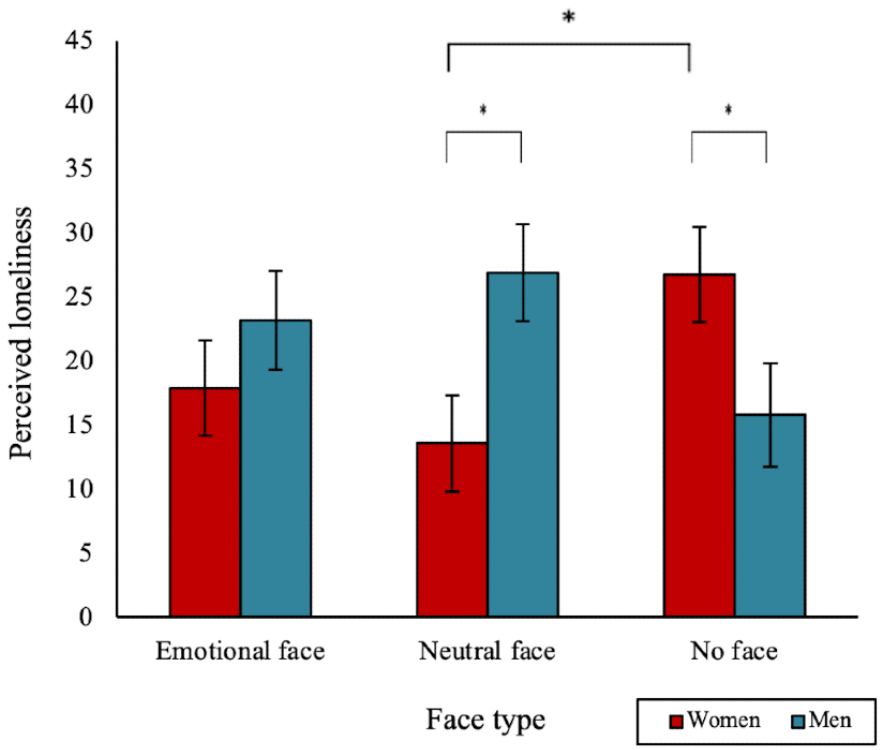
A manipulation check revealed that participants perceived the emotional face and voice conditions as more emotionally expressive. There were significant main effects of face type ($F_{2,183}$=6.97; P=.001) and voice type ($F_{1,183}$=3.94; P=.049) on perceived emotional expressiveness of the digital human. An emotional face was rated as significantly more emotionally expressive than a neutral face (mean 53.02, SD 27.33, vs mean 40.14, SD 23.06; P=.01). No face was rated as more emotionally expressive than a neutral face (mean 55.17, SD 24.90; P=.002). In terms of voice type, emotional voice was rated as significantly more emotionally expressive than neutral voice (mean 52.79, SD 24.73, vs mean 45.91, SD 26.71; P=.049). There were no other significant main, 2-way, or 3-way interaction effects (all F values <1.60).

At times, errors with the conversation engine meant that the digital human occasionally interrupted the participants during their speech. However, interruptions were equally likely across the experimental conditions ($F_{5,179}$=1.11; P=.36).

Perceived Loneliness (T2)

Analyses revealed gender differences in the effect of a digital human's face type on perceived loneliness. There was a significant 2-way interaction effect between face type and user gender ($F_{2,182}$=4.95; P=.008; Figure 4). A digital human with no face was associated with significantly less loneliness in men than in women (mean 15.76, SD 19.07, vs mean 26.71, SD 24.86; P=.047). A neutral face digital human was associated with significantly less loneliness in women than in men (mean 13.55, SD 17.11, vs mean 26.88, SD 26.64; P=.01). A neutral face digital human was also associated with significantly less loneliness in women than a no-face digital human (P=.04). There were no other significant main, 2-way, or 3-way interaction effects (all F values <1.68).

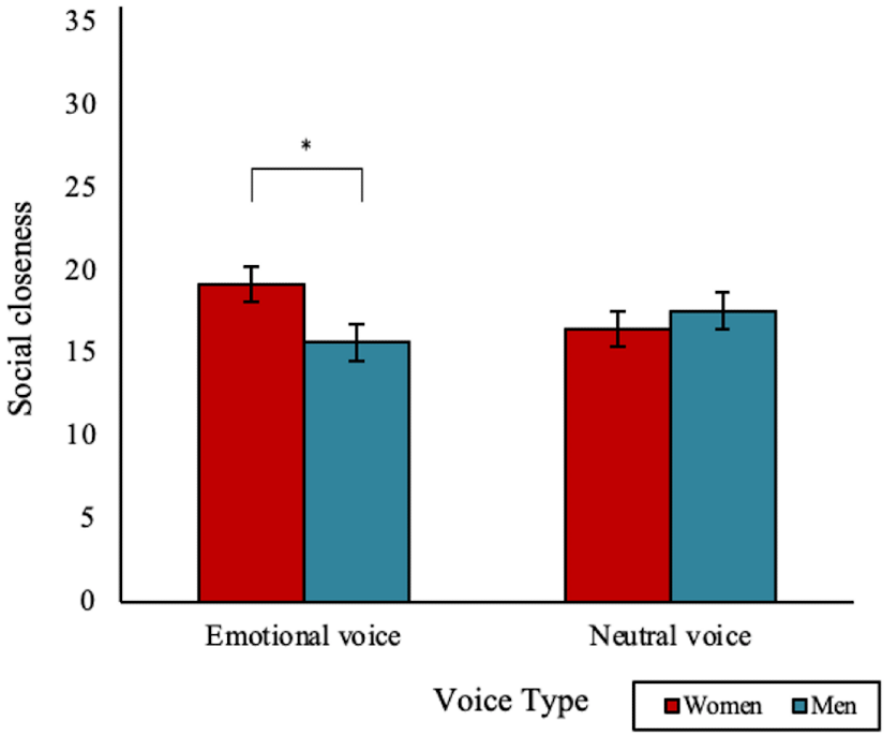
Figure 4. A significant two-way interaction effect of face type and gender on perceived loneliness after the digital human conversation (*P<.05). Bars indicate the SE.



Social Closeness

An emotional voice increased feelings of closeness toward a digital human but only in women. There was a significant 2-way interaction effect between voice type and gender ($F_{1,181}$=4.17; *P*=.04; Figure 5). Women reported significantly greater closeness with an emotional voice digital human than men did (mean 19.18, SD 7.21, vs mean 15.68, SD 7.53; *P*=.02). There was a trend toward women reporting greater closeness with an emotional voice digital human compared with a neutral voice digital human (mean 16.45, SD 7.06; *P*=.07). There were no other significant main, 2-way, or 3-way interaction effects (all *F* values <1.06).

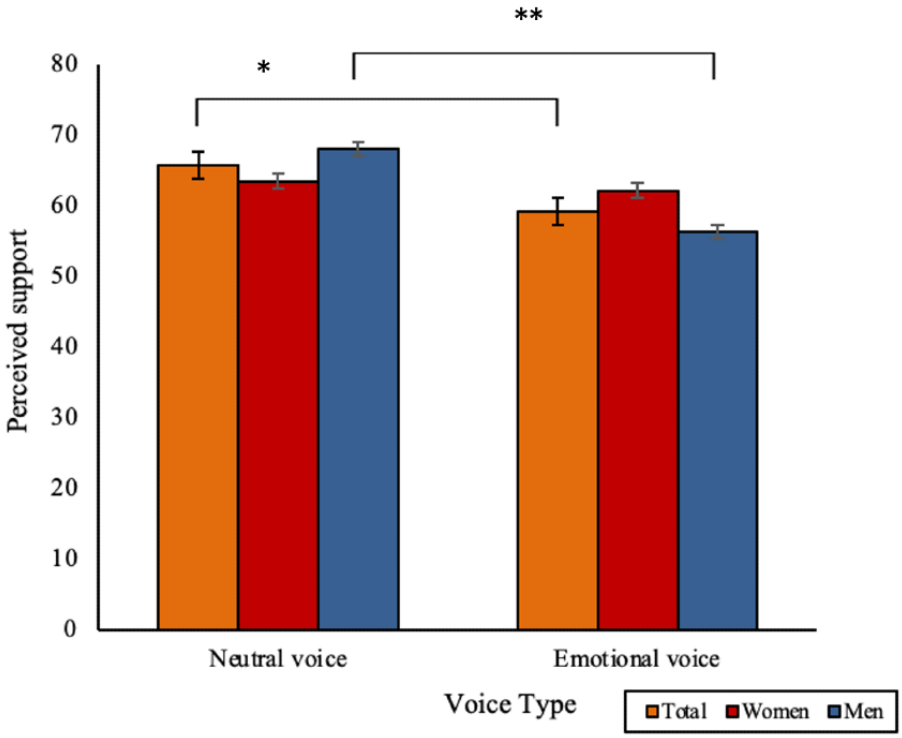
Figure 5. A significant two-way interaction effect of voice type and gender on social closeness with a digital human after the conversation (*P<.05). Bars indicate the SE.



Perceived Support (T2)

A neutral voice was associated with significantly greater feelings of support across men and women. Analyses revealed a significant main effect of voice type on perceived support ($F_{1,182}$=5.77; *P*=.02). A neutral voice digital human was associated with significantly higher ratings of perceived support than an emotional voice digital human (mean 65.77, SD 19.09, vs mean 59.23, SD 19.01; *P*=.02; Figure 6).

Figure 6. A significant main effect of digital human voice type on perceived support after the digital human conversation (*P<.05 and **P<.01). Bars indicate the SE.



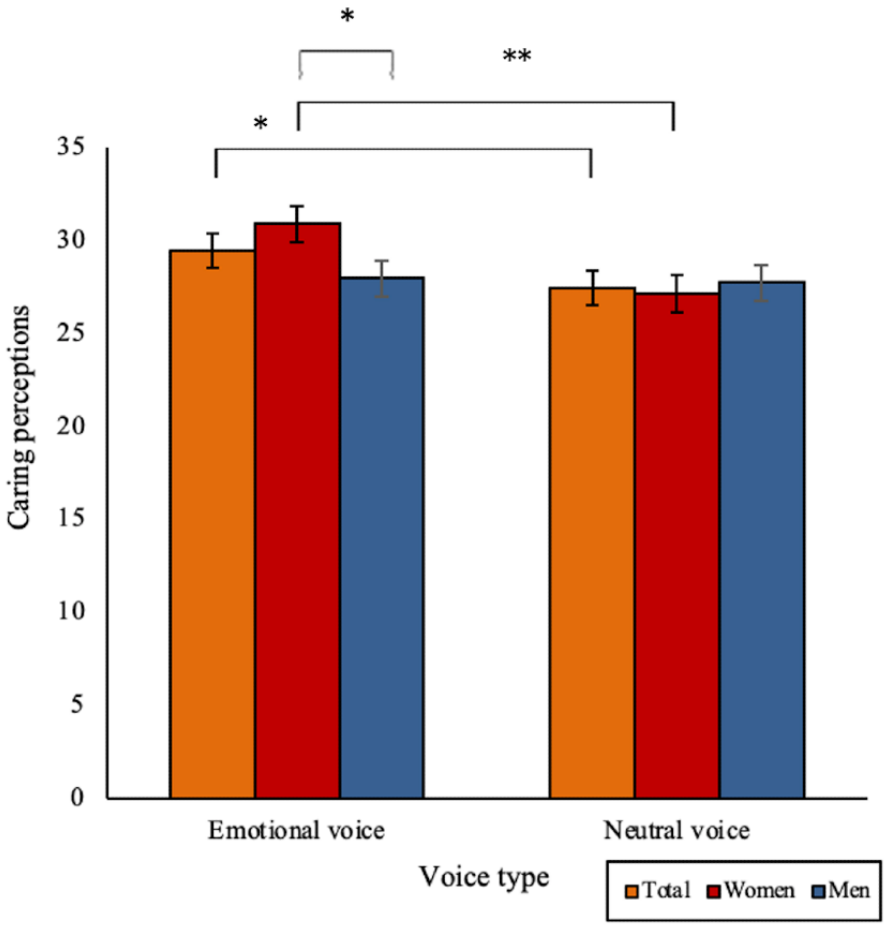
Analyses also revealed that there may be gender differences in how supported people felt in response to emotional and neutral voices in digital humans. There was a trend toward a significant 2-way interaction effect between voice type and gender ($F_{1,182}$=3.66; P=.06). Men felt significantly more supported with a neutral voice digital human than with one with an emotional voice (mean 68.03, SD 18.49, vs mean 56.29, SD 18.38; P=.002), whereas women felt more supported with an emotional voice digital human than men (mean 62.18, SD 19.25, vs mean 56.29, SD 18.38), however, this was a trend toward significance (P=.10). There were no other significant main, 2-way, or 3-way interaction effects (all F values <2.33).

Caring Perceptions

An emotional voice was found to increase caring perceptions in both men and women. Analyses revealed a significant main effect of voice type on caring perceptions ($F_{1,183}$=4.26; P=.04). An emotional voice digital human was rated as significantly more caring than a neutral voice digital human (mean 29.45, SD 6.56, vs mean 27.42, SD 6.94; P=.04).

There was a trend toward a significant 2-way interaction effect between voice type and gender ($F_{1,183}$=3.31; P=.07; Figure 7). Women rated the emotional voice digital human as significantly more caring than men did (mean 30.88, SD 5.82, vs mean 27.96, SD 6.99; P=.03). Women also rated the emotional voice digital human as significantly more caring than a neutral voice digital human (mean 27.14, SD 6.83; P=.006). There were no other significant main, 2-way, or 3-way interaction effects (all F values <1.55).

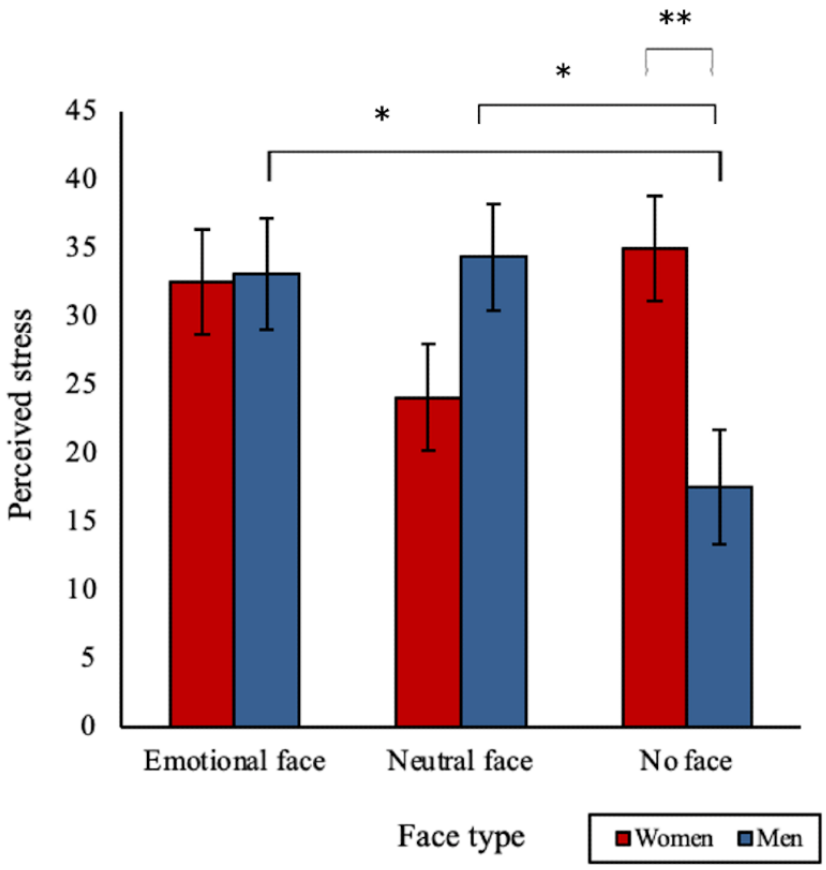
Figure 7. A significant main effect of voice type on caring perceptions, with a trend for an interaction with gender (*P<.05 and **P<.01). Bars indicate the SE.



Perceived Stress (T2)

Gender was found to impact the effect of a digital human's face type on perceived stress. A significant 2-way interaction effect was found between face type and gender ($F_{2,182}$=6.01, P=.003; Figure 8). A nonface digital human was associated with significantly lower perceived stress in men than a neutral face (mean 17.55, SD 15.87, vs mean 34.36, SD 22.53; P=.01) and an emotional face digital human (mean 33.13, SD 19.09; P=.02), whereas women reported significantly greater stress with a no-face digital human than men (mean 35.00, SD 23.90; P=.003). Women also reported lower stress with a neutral face digital human than men; however, this was a trend toward significance (mean 24.09, SD 23.67; P=.07). There were no other significant main, 2-way, or 3-way interaction effects (all F values <2.68).

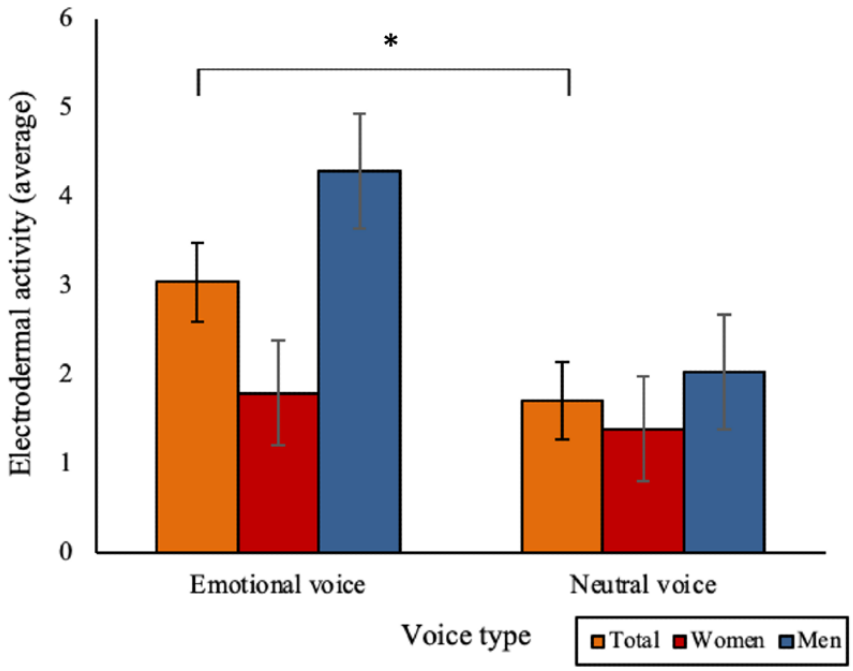
Figure 8. A significant two-way interaction effect of face type and gender on perceived stress after the digital human interaction (*P<.05 and **P<.01). Bars indicate the SE.



Physiological Outcomes

There were no significant main, 2-way, or 3-way interaction effects of face type, voice type, or gender on average skin temperature (all F values <0.89) or heart rate (all F values <0.98) over the conversation. However, there were significant main effects of voice type ($F_{1,158}$=4.72; P=.03) and gender ($F_{1,158}$=6.54; P=.01) on average electrodermal activity. An emotional voice was associated with significantly greater electrodermal activity over the conversation than a neutral voice (mean 3.07, SD 5.38, vs mean 1.71, SD 2.13; P=.03; Figure 9). In addition, men experienced higher electrodermal activity during the interaction than women (mean 3.19, SD 5.69, vs mean 1.59, SD 2.04; P=.01). There were no other significant main, 2-way, or 3-way interaction effects (all F values <2.38).

Figure 9. Significant main effects of voice type and gender on electrodermal activity during the digital human interaction (*P<.05). Bars indicate the SE.



Discussion

Principal Findings

This is the first study to investigate whether emotional expressiveness in a female digital human and user gender interact to affect loneliness, social, and physiological outcomes following a self-disclosure conversation.

Several main effects were found for the impact of emotional expressiveness in a female digital human's voice on physiological arousal and social outcomes. Perceived social support was significantly higher for participants after interacting with a neutral voice digital human than after interacting with an emotional voice digital human. However, an emotional voice digital human was perceived as more caring overall than one with a neutral voice. Average electrodermal activity was higher during interactions with an emotional voice digital human than with a neutral voice one. These findings suggest that emotional expression in a digital human's voice can increase physiological arousal and perceptions of caring yet reduce feelings of being supported across both genders.

Gender *did* impact the effect of emotional expressiveness on loneliness, physiological arousal, and social outcomes. Women benefited more from a female digital human with emotional expressions in the voice and with a face. For women, emotional expression in the voice increased closeness with the digital human, caring perceptions, and perceived social support. In addition, for women, the presence of a neutral face was associated with reduced loneliness and subjective stress than a digital human with no face.

In contrast, men had better outcomes with a female digital human without emotional expression in the voice and with no face (ie, a black screen). For men, a neutral voice digital human was associated with increased closeness to the digital human, caring perceptions, and perceived social support. A digital human without a face (ie, a black screen) was associated with less loneliness and lower subjective stress in men.

The potential reasons for these effects and how these findings relate to other research are discussed in subsequent sections.

Main Effects

Emotional expression in the voice increased physiological arousal and perceptions of care yet reduced perceptions of support. The reason for why people felt less supported may be because the emotions were focused on the digital humans' own words, as opposed to varying in response to the emotions expressed by the user. It is likely that the users expected the digital human to respond emotionally to their personal stories. Indeed, prior research has found that a computer agent that consistently delivered other focused emotions (eg, compassion) in response to users' behavior in a game was perceived as more caring and was associated with greater feelings of support than a computer agent that delivered self-focused emotion (eg, joy at its own success in the game) [55].

The emotional voice was associated with greater physiological arousal than a neutral voice. High electrodermal activity could indicate either positively or negatively valanced arousal (eg, excitement or fear) [69]. This result is consistent with previous research showing that electrodermal activity increases when listening to emotionally valanced sounds both pleasant and

unpleasant, although the greatest increases occur in response to pleasant sounds [69].

Men were found to have greater electrodermal activity than women during the interaction with the digital human. This may be explained by prior research showing that men have different brain responses to expressed emotions than women, along with different responses to viewing a female face [70,71]. For example, men and women have been shown to differentially activate brain regions in response to observing contempt and disgust facial expressions (eg, the medial frontal gyrus) [69]. This differential activation of brain regions may have downstream physical effects on the body, influencing outcomes such as electrodermal activity.

Gender Differences

The finding that women felt closer to a digital human with an emotional voice but men felt closer to a digital human with a neutral voice is in keeping with prior literature. Emotional expressiveness is an important part of close female relationships [41,43,44], but it is a less important part of close male relationships. For men, shared activities, interests, and ideas are more important aspects of close relationships [41,44].

Similar results have been found in research examining the effect of emotional expressiveness on closeness with other types of artificial agents, including robots. Women felt closer to social robots that expressed emotions during interactions, whereas men felt closer to robots that expressed fewer emotions [54]. In research with computer agents, affective support in language was associated with greater rapport in girls, whereas for boys, an agent with only task-oriented language was associated with higher rapport [72]. The results of this study build upon the literature by showing that emotional expressiveness by computer agents can foster closer relationships for women, especially when emotions are expressed in the voice.

The gendered effects of the emotionality of the voice on perceived closeness may have also had an impact on perceived support. Women felt more supported by the emotional voice and men felt more supported by the neutral voice. This aligns with previous research showing that perceptions of social support are related to perceived closeness toward the support partner in human relationships [39].

This research also found gender effects on the impact of the digital human's face type on perceived loneliness and stress. Men felt least lonely and stressed after interacting with a no-face digital human. In contrast, women felt the least lonely and stressed with a neutral face digital human. These results were unexpected; we predicted that women would experience the least amount of loneliness and stress with an emotional face and voice digital human. This was based on the existing literature showing lower loneliness and stress after interactions with emotionally expressive computer agents [73,74]. However, the effects of emotional expression on the face and in the voice were not compared in these studies.

This is the first study to investigate the effects of emotional expression in the voice and on the face separately in computer agents. It is possible that the effects of emotional expression are stronger when expressed in the voice than on the face. This is supported by findings that have shown that voice-only communication can increase perceptions of empathic accuracy [42]. However, as the emotional faces were very subtle in their expressions and the neutral face included humanlike movements that could have inadvertently portrayed interest (eg, raising eyebrows with speech), it is possible that the emotional and neutral face conclusions are both more about the presence of a face. This study leaves open the question of the degree of emotional expression; it is possible that different results may have been achieved with more pronounced emotional expressions on the face.

It is unclear why a neutral face was found to be more helpful for loneliness and stress in women than an emotional face, given that a manipulation check revealed that the emotional face was perceived as more emotionally expressive. It is possible that either the emotional expressions were not strong enough or the types of emotions that the digital human expressed in this study influenced outcomes. The digital human expressed a combination of self- (eg, joy and sadness) and other focused (eg, compassion and concern) facial emotions during the interaction. An interaction involving only other focused emotions could have more accurately mimicked the role of a supportive peer and had stronger effects on loneliness and stress.

Other research has shown that smiling is associated with better perceptions of a health care robot than not smiling [75]. In this study, the emotional face digital human did not smile all the time, and instead had a slight frown when expressing concern and looked sad when expressing its own sad experiences. The presence of one or both of these negative emotions might have negatively affected outcomes.

This study adds to the literature by showing that emotional expressiveness (a behavior that is important in the development of close relationships in women but not men) shows similar effects in relationships with a computer agent. In addition, the results of this study show that loneliness, stress, and social outcomes that have been associated with close human relationships can also be influenced by conversations with a computer agent. The results provide some supportive evidence for the CASA paradigm [46], which suggests that people engage in the same social behaviors toward computers that elicit social cues as they do toward other people. It makes sense that behaviors that help to build closeness in human relationships (eg, emotional expressiveness) have similar effects in relationships with a computer agent. However, technological limitations in responding may have affected cues being entirely the same as for real human conversations, as discussed in the *Limitations* section.

Design Recommendations

Overall, the results of this study provide several design recommendations for digital humans intended for the role of a supportive peer. First, when designing a supportive peer for female users, incorporating an emotional voice may be important for promoting a closer relationship and feelings of support. For women, including an animated face was also important for reducing loneliness and perceived stress.

In contrast, when designing a supportive peer for men, a neutral voice may help to promote a closer relationship and increase feelings of support. For men, no face (ie, a black screen with a voice) was associated with less loneliness and stress than a female face with some or no emotional expression. However, it is possible that different results may have been found for men had they interacted with a male digital human.

It is important to note that these recommendations were based on results obtained with a predominantly young adult, New Zealand student sample and with a young female digital human. However, previous research has also shown gender differences in the effects of different closeness-building behaviors with other age groups in different geographic locations [41,43,44].

These findings may not generalize to all cultures. It is possible that people's preferences regarding different aspects of emotional expressiveness (eg, type and strength) may shift at the intersection of gender and other demographic characteristics, such as culture. Prior research has shown cultural differences in targets for emotional expressiveness, including how emotions are expressed and what types are the most socially desirable to portray [76,77]. For example, in East Asian cultures, it is more socially desirable to express low arousal positive affect (eg, calm), and negative emotions are less often expressed directly [77]. In this study, there were 41% (80/195) Asian participants and 34.4% (67/195) New Zealand Europeans.

In contrast, in North American White culture, high arousal positive affect (eg, excitement) is the most socially desirable to express [77]. As a result, culture may affect how a digital human's emotional expressiveness is perceived by people and, subsequently, the effect it has on loneliness, social, and physiological outcomes. It is possible that emotional expressiveness in a digital human may need to be tailored to the culture of its intended users alongside their gender. The results may have been different if the study was conducted in the United States.

Limitations

This study had several limitations that could affect the generalizability of the results. First, the digital human conversation engine experienced occasional errors with utterance detection and speech-to-text translation during its interactions with people. These errors manifested as behaviors such as interrupting people or asking people to repeat themselves, which could have increased negative mood in some participants [68]. However, the errors were not more or less frequent in any of the experimental conditions.

Another possible limitation is that the study was conducted in a community sample predominantly comprised of young adult students. It is unclear whether the findings would apply to other groups who might benefit from a social support intervention with a digital human, such as older adults (who could have lower digital literacy and more challenges with using a digital human program) or particular patient samples (who may be considerably more stressed at baseline). In addition, this study focused on gender differences between men and women; therefore, it is unknown how gender-diverse persons might respond to emotional expressiveness in a digital human, including the effects on their loneliness, social, and physiological outcomes.

Another consideration is that other features of the digital human could have inadvertently influenced people's responses to it. For example, research has shown that the proportion of a computer agent's facial features may influence personality perceptions [78]. A narrower face and smaller eyes have been associated with perceptions of greater aggression in computer agents [78]. Both narrow and wide faces have been associated with lower perceptions of trustworthiness than a standard width face, and smaller eyes have been rated as less trustworthy than medium-sized eyes in a computer agent [78]. The face design of the digital human in this study was consistent across the neutral and emotional face conditions; however, each of these components of the digital human's face could have contributed in part to how participants evaluated the digital human. It is also possible that perceived attractiveness and eeriness of the digital human (ie, the uncanny valley effect [79]), as well as prior experiences with digital humans could have influenced the responses. However, because participants were randomized into groups, these factors should not confound the results.

An important consideration is that demographic matching (eg, gender and ethnicity) between the participant and the digital human could have influenced the social responses. The digital human in this study was female, and different results may have been found if a male digital human had been used. In prior research, gender matching has been shown to influence preferences toward computer agents [80], and ethnicity matching has increased perceptions of trustworthiness [81], social presence [82], and credibility toward computer agents [83]. Ethnicity matching has also been shown to increase closeness, disclosure comfort, and future relationship expectations with computer agents [84].

In this study, the digital human was of mixed ethnicity (New Zealand Māori, New Zealand European, Indian, and European), and the participants were predominantly New Zealand European or Asian, but all resided in New Zealand. There was no significant difference between groups in the proportions of participants of each ethnicity; therefore, this would not have affected our main results. Nevertheless, across the entire sample, Asian ethnicity was related to greater perceived closeness with the digital human (P=.01). This may be because of the tendency for Asian cultures to be more accepting of robotic technologies in general [85]. Matching the user's accent and conversational style has also been shown to increase acceptability and trust with robots and computer agents [86-90]. In this study, the digital human spoke in a local New Zealand accent; however, she did not match the conversational style of participants. We do not have data on participants' accents from this study.

The effects of demographic and conversational style matching may be because of the similarity-attraction hypothesis, which predicts that people like others more who are more similar to themselves [91]. This social phenomenon has been shown to apply to relationships with computers [46]. Matching the digital human's demographic characteristics to each participant was not feasible in this study; however, future research could investigate how this might affect the results. The

similarity-attraction hypothesis may also explain some gender effects, as women are generally more emotionally expressive than men.

Finally, the expressions in the emotional face condition may not have been expressive enough to elicit changes in outcomes for users. It is possible that more pronounced emotional expressions may have resulted in different findings. In addition, humanlike behaviors performed by the neutral face condition could have been interpreted as showing interest, which may have influenced the results (eg, head movements while listening and eyebrow raises while speaking).

Future Research

The results of this study indicate several important areas for future research. First, experimental research is needed to investigate the generalizability of the results at the intersection of gender and other demographic or health characteristics (eg, ethnicity, age, and disability). In addition, this study did not examine the effect of emotional expressiveness in a digital human on individuals who are gender diverse, and this topic should be explored in future research.

Second, it is possible that the results may have been affected by the degree of emotional expression in the digital human's face and the types of emotions that the digital human expressed during the interaction (ie, a mixture of self- and other focused emotions and positive and negative emotions). The expression of these types of emotions has been shown to influence perceptions of computer agents and robots in prior research [55,75]. Future studies could investigate whether expressing more pronounced emotions on the face improves outcomes and evaluate the effect of other focused emotional expression in a digital human (eg, compassion and concern) on loneliness, social, and physiological outcomes. It is likely that this style of emotional expression would better mimic a therapeutic or supportive interaction and may result in greater improvements in outcomes than a digital human that expresses mixed or self-focused emotion. Research could also compare the effects of expressing positive, negative, and a mixture of these expressions.

This study identified emotional expressiveness as an important design feature that should be included in female digital human support persons for women. However, there are likely many other digital human behaviors that could contribute to reductions in loneliness, social, and physiological outcomes for people [40]. This study found that men experienced better outcomes with a female digital human that had less emotional expressiveness in the voice and had no face. However, this finding should be further investigated as men (and women) could have responded differently to a male digital human.

Conclusions

Overall, emotional expressiveness in a female digital human's voice was associated with perceiving her as more caring and with experiencing greater physiological arousal during a self-disclosure conversation. However, emotion in the voice was associated with lower feelings of support after the interaction. Gender was an important moderator of these effects. The findings indicate that when designing a female digital human support person for women, an emotional voice and the presence of a face may be important features to include. In contrast, when designing a female digital support person for men, a neutral voice and no face may result in better outcomes than a female with emotional expressiveness. The results provide support for the CASA paradigm and may inform the design of computer agent support persons.

Acknowledgments

The authors wish to thank the engineering team at Soul Machines Ltd for their assistance in building the digital human for this experiment. The authors would also like to thank Andy Ha for assistance with sensor data processing.

Conflicts of Interest

MS is the chief executive officer of Soul Machines (an artificial intelligence company), which supports KL with a PhD stipend, contracts EB for consultancy work, and hired XZ as an engineer at the time of the research.

Multimedia Appendix 1
CONSORT e-HEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 397 KB - jmir_v23i11e30624_app1.pdf]

References

1. Bu F, Steptoe A, Fancourt D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health 2020 Sep;186:31-34 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.036] [Medline: 32768621]
2. Perlman D, Peplau LA. Toward a social psychology of loneliness. In: Duck S, Gilmour R, editors. Personal Relationships in Disorder. Cambridge, Massachusetts: Academic Press; 1981:31-56.
3. Cacioppo S, Capitanio JP, Cacioppo JT. Toward a neurology of loneliness. Psychol Bull 2014 Nov;140(6):1464-1504 [FREE Full text] [doi: 10.1037/a0037618] [Medline: 25222636]
4. Masi CM, Chen H, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug 17;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]

5. Jones WH, Moore TL. Loneliness and social support. J Soc Behav Personal 1987;2(2):145-156 [FREE Full text]
6. Groarke JM, Berry E, Graham-Wisener L, McKenna-Plumley PE, McGlinchey E, Armour C. Loneliness in the UK during the COVID-19 pandemic: cross-sectional results from the COVID-19 Psychological Wellbeing Study. PLoS One 2020 Sep 24;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
7. Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015 May;105(5):1013-1019. [doi: 10.2105/AJPH.2014.302427] [Medline: 25790413]
8. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
9. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
10. Lee SL, Pearce E, Ajnakina O, Johnson S, Lewis G, Mann F, et al. The association between loneliness and depressive symptoms among adults aged 50 years and older: a 12-year population-based cohort study. Lancet Psychiatry 2021 Jan;8(1):48-57 [FREE Full text] [doi: 10.1016/S2215-0366(20)30383-7] [Medline: 33181096]
11. Beutel ME, Klein EM, Brähler E, Reiner I, Jünger C, Michal M, et al. Loneliness in the general population: prevalence, determinants and relations to mental health. BMC Psychiatry 2017;17(1):97 [FREE Full text] [doi: 10.1186/s12888-017-1262-x] [Medline: 28320380]
12. Bu F, Steptoe A, Fancourt D. Loneliness during a strict lockdown: trajectories and predictors during the COVID-19 pandemic in 38,217 United Kingdom adults. Soc Sci Med 2020;265:113521 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113521] [Medline: 33257177]
13. Uchino BN, Cacioppo JT, Kiecolt-Glaser JK. The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. Psychol Bull 1996;119(3):488-531. [doi: 10.1037/0033-2909.119.3.488] [Medline: 8668748]
14. Hogan BE, Linden W, Najarian B. Social support interventions: do they work? Clin Psychol Rev 2002;22(3):381-440. [doi: 10.1016/s0272-7358(01)00102-7]
15. Lepore SJ. Problems and prospects for the social support-reactivity hypothesis. Ann Behav Med 1998;20(4):257-269. [doi: 10.1007/BF02886375] [Medline: 10234419]
16. Ditzen B, Heinrichs M. Psychobiology of social support: the social dimension of stress buffering. Restor Neurol Neurosci 2014;32(1):149-162. [doi: 10.3233/RNN-139008] [Medline: 23603443]
17. Kiecolt-Glaser JK, Gouin J, Hantsoo L. Close relationships, inflammation, and health. Neurosci Biobehav Rev 2010;35(1):33-38 [FREE Full text] [doi: 10.1016/j.neubiorev.2009.09.003] [Medline: 19751761]
18. Heinrichs M, Baumgartner T, Kirschbaum C, Ehlert U. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biolog Psychiatry 2003;54(12):1389-1398. [doi: 10.1016/S0006-3223(03)00465-7]
19. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
20. Cohen S, Janicki-Deverts D, Doyle WJ, Miller GE, Frank E, Rabin BS, et al. Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. Proc Natl Acad Sci 2012;109(16):5995-5999 [FREE Full text] [doi: 10.1073/pnas.1118355109] [Medline: 22474371]
21. Glaser R, Kiecolt-Glaser J. How stress damages immune system and health. Discov Med 2005;5(26):165-169 [FREE Full text] [Medline: 20704904]
22. Fricchione G. Compassion and Healing in Medicine and Society: On the Nature and Use of Attachment Solutions to Separation Challenges. Baltimore, MD: Johns Hopkins University Press; 2011:1-552.
23. Hegel F, Muhl C, Wrede B, Hielscher-Fastabend M, Sagerer G. Understanding social robots. In: Proceedings of the Second International Conference on Advances in Computer-Human Interactions. 2009 Presented at: Second International Conference on Advances in Computer-Human Interactions; Feb 1-7, 2009; Cancun, Mexico p. 169-174. [doi: 10.1109/achi.2009.51]
24. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
25. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55(2):569-574 [FREE Full text] [doi: 10.3233/JAD-160703] [Medline: 27716673]
26. Robinson H, MacDonald B, Broadbent E. Physiological effects of a companion robot on blood pressure of older people in a residential care facility: a pilot study. Australas J Ageing 2015;34(1):27-32. [doi: 10.1111/ajag.12099] [Medline: 24373064]
27. Moyle W, Jones C, Cooke M, O'Dwyer S, Sung B, Drummond S. Connecting the person with dementia and family: a feasibility study of a telepresence robot. BMC Geriatr 2014;14(1):7 [FREE Full text] [doi: 10.1186/1471-2318-14-7] [Medline: 24456417]
28. Niemelä M, van Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards ethical guidelines of using telepresence robots in residential care. Int J of Soc Robotics 2019;13(3):431-439. [doi: 10.1007/s12369-019-00529-8]

29. Powers A, Kiesler S, Fussell S, Torrey C. Comparing a computer agent with a humanoid robot. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction. 2007 Presented at: International Conference on Human Robot Interaction; March 10 - 12, 2007; Arlington, VA, USA p. 145-152. [doi: 10.1145/1228716.1228736]
30. Cassell J, Sullivan J, Prevost S, Churchill EF. Embodied Conversational Agents. Cambridge, MA: MIT Press; 2000.
31. Gasteiger N, Loveys K, Law M, Broadbent E. Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. Clin Interv Aging 2021;Volume 16:941-971. [doi: 10.2147/cia.s282709]
32. Bott N, Wexler S, Drury L, Pollak C, Wang V, Scher K, et al. A protocol-driven, bedside digital conversational agent to support nurse teams and mitigate risks of hospitalization in older adults: case control pre-post study. J Med Internet Res 2019;21(10):e13440 [FREE Full text] [doi: 10.2196/13440] [Medline: 31625949]
33. Ring L, Shi L, Totzke K, Bickmore T. Social support agents for older adults: longitudinal affective computing in the home. J Multimodal User Interfaces 2014;9(1):79-88. [doi: 10.1007/s12193-014-0157-0]
34. Vardoulakis LP, Ring L, Barry B, Sidner CL, Bickmore T. Designing relational agents as long term social companions for older adults. In: Proceedings of the International Conference on Intelligent Virtual Agents. 2012 Presented at: International Conference on Intelligent Virtual Agents; September 12-14, 2012; Santa Cruz, CA, USA p. 289-302. [doi: 10.1007/978-3-642-33197-8_30]
35. Kim J, Kim Y, Kim B, Yun S, Kim M, Lee J. Can a machine tend to teenagers' emotional needs?: a study with conversational agents. In: Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal, QC, Canada p. 1-6. [doi: 10.1145/3170427.3188548]
36. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
37. Sagar M, Bullivant D, Robertson P, Efimov O, Jawed K, Kalarot R, et al. A neurobehavioural framework for autonomous animation of virtual human faces. In: Proceedings of the SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence. 2014 Presented at: SA'14: SIGGRAPH Asia 2014; December 3 - 6, 2014; Shenzhen China. [doi: 10.1145/2668956.2668960]
38. Sagar M, Seymour M, Henderson A. Creating connection with autonomous facial animation. Commun ACM 2016;59(12):82-91. [doi: 10.1145/2950041]
39. Rini C, Schetter CD, Hobel CJ, Glynn LM, Sandman CA. Effective social support: antecedents and consequences of partner support during pregnancy. Pers Relat 2006;13(2):207-229. [doi: 10.1111/j.1475-6811.2006.00114.x]
40. Loveys K, Sebaratnam G, Sagar M, Broadbent E. The effect of design features on relationship quality with embodied conversational agents: a systematic review. Int J of Soc Robotics 2020;12(6):1293-1312 [FREE Full text] [doi: 10.1007/s12369-020-00680-7]
41. Parks MR, Floyd K. Meanings for closeness and intimacy in friendship. J Soc Pers Relat 2016;13(1):85-107. [doi: 10.1177/0265407596131005]
42. Burleson BR. The experience and effects of emotional support: what the study of cultural and gender differences can tell us about close relationships, emotion, and interpersonal communication. Pers Relat 2003;10(1):1-23. [doi: 10.1111/1475-6811.00033]
43. Floyd K. Gender and closeness among friends and siblings. J Psychol 1995;129(2):193-202 [FREE Full text] [doi: 10.1080/00223980.1995.9914958] [Medline: 7760294]
44. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Soc Isssues 2000;56(1):81-103. [doi: 10.1111/0022-4537.00153]
45. Kraus MW. Voice-only communication enhances empathic accuracy. Am Psychol 2017;72(7):644-654. [doi: 10.1037/amp0000147] [Medline: 29016168]
46. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge, UK: Cambridge University Press; 1997:1-324.
47. Nass C, Lee KM. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. J Exp Psychol Appl 2001;7(3):171-181. [doi: 10.1037/1076-898x.7.3.171]
48. Nass C, Moon Y, Green N. Are machines gender neutral? Gender-stereotypic responses to computers with voices. J Appl Social Pyschol 1997;27(10):864-876. [doi: 10.1111/j.1559-1816.1997.tb00275.x]
49. Nass C, Fogg B, Moon Y. Can computers be teammates? Int J Hum Comput Stud 1996;45(6):669-678. [doi: 10.1006/ijhc.1996.0073]
50. Lee S, Ratan R, Park T. The voice makes the car: enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. Multimodal Technol Interact 2019;3(1):20. [doi: 10.3390/mti3010020]
51. Moon Y, Kim KJ, Shin DH. Voices of the Internet of Things: an exploration of multiple voice effects in smart homes. In: Proceedings of the International Conference on Distributed, Ambient, and Pervasive Interactions. 2016 Presented at: International Conference on Distributed, Ambient, and Pervasive Interactions; July 17-22, 2016; Toronto, Canada p. 270-278. [doi: 10.1007/978-3-319-39862-4_25]

52. Gaia AC. Understanding emotional intimacy: a review of conceptualization, assessment and the role of gender. Int Soc Sci Rev 2002;77(3/4):151-170 [FREE Full text]
53. MacLean PD. Brain evolution relating to family, play, and the separation call. Arch Gen Psychiatry 1985;42(4):405-417. [doi: 10.1001/archpsyc.1985.01790270095011] [Medline: 3977559]
54. Kory-Westlund JM. Relational AI: creating long-term interpersonal interaction, rapport, and relationships with social robots. PhD Thesis, Massachusetts Institute of Technology. 2019. URL: https://hdl.handle.net/1721.1/123627 [accessed 2021-11-05]
55. Brave S, Nass C, Hutchinson K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Hum-Comput Stud 2005 Feb;62(2):161-178. [doi: 10.1016/j.ijhcs.2004.11.002]
56. Maldonado H, Lee JR, Brave S, Nass C, Nakajima H, Yamada R, et al. We learn better together: enhancing eLearning with emotional characters. In: Proceedings of the Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!. 2005 Presented at: Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!; May 30 - June 4, 2005; Taipei Taiwan p. 408-417. [doi: 10.3115/1149293.1149347]
57. Koda T, Maes P. Agents with faces: the effect of personification. In: Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA. 1996 Presented at: 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA; Nov. 11-14, 1996; Tsukuba, Japan p. 189-194. [doi: 10.1109/roman.1996.568812]
58. Aron A, Aron EN, Smollan D. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. J Personal Soc Psychol 1992;63(4):596-612. [doi: 10.1037/0022-3514.63.4.596]
59. Popovic M, Milne D, Barrett P. The scale of perceived interpersonal closeness (PICS). Clin Psychol Psychother 2003 Sep;10(5):286-301. [doi: 10.1002/cpp.375]
60. Sedikides C, Campbell WK, Campbell GD, Elliot AJ. The relationship closeness induction task. Represent Res Soc Psychol 1999;23:1-4 [FREE Full text]
61. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010;11(1):32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
62. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [doi: 10.1037/0022-3514.39.3.472]
63. Aitken RC. Measurement of feelings using visual analogue scales. Proc R Soc Med 1969 Oct;62(10):989-993 [FREE Full text] [Medline: 4899510]
64. Robinson H, Ravikulan A, Nater UM, Skoluda N, Jarrett P, Broadbent E. The role of social closeness during tape stripping to facilitate skin barrier recovery: preliminary findings. Health Psychol 2017 Jul;36(7):619-629. [doi: 10.1037/hea0000492] [Medline: 28277705]
65. Law M, Jarrett P, Nater UM, Skoluda N, Broadbent E. The effects of environmental enrichment on skin barrier recovery in humans: a randomised trial. Sci Rep 2020 Jun 17;10(1):9829 [FREE Full text] [doi: 10.1038/s41598-020-66687-2] [Medline: 32555211]
66. Lazarus RS, Speisman JC, Mordkoff AM. The relationship between autonomic indicators of psychological stress: heart rate and skin conductance. Psychosom Med 1963;25(1):19-30. [doi: 10.1097/00006842-196301000-00004]
67. Rimm-Kaufman SE, Kagan J. The psychological significance of changes in skin temperature. Motiv Emot 1996 Mar;20(1):63-78. [doi: 10.1007/bf02251007]
68. Loveys K, Sagar M, Broadbent E. The effect of multimodal emotional expression on responses to a digital human during a self-disclosure conversation: a computational analysis of user language. J Med Syst 2020 Jul 22;44(9):143. [doi: 10.1007/s10916-020-01624-4] [Medline: 32700060]
69. Bradley MM, Lang PJ. Affective reactions to acoustic stimuli. Psychophysiology 2000 Mar;37(2):204-215. [doi: 10.1111/1469-8986.3720204]
70. Aleman A, Swart M. Sex differences in neural activation to facial expressions denoting contempt and disgust. PLoS One 2008 Nov 5;3(11):e3622 [FREE Full text] [doi: 10.1371/journal.pone.0003622] [Medline: 18985147]
71. Kveraga K, De Vito D, Cushing C, Im HY, Albohn DN, Adams RB. Spatial and feature-based attention to expressive faces. Exp Brain Res 2019 Apr 25;237(4):967-975 [FREE Full text] [doi: 10.1007/s00221-019-05472-8] [Medline: 30683957]
72. Burleson W, Picard RW. Gender-specific approaches to developing emotionally intelligent learning companions. IEEE Intell Syst 2007 Jul;22(4):62-69. [doi: 10.1109/mis.2007.69]
73. Ring L, Barry B, Totzke K, Bickmore T. Addressing loneliness and isolation in older adults: proactive affective agents provide better support. In: Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: Humaine Association Conference on Affective Computing and Intelligent Interaction; Sept. 2-5, 2013; Geneva, Switzerland p. 61-66. [doi: 10.1109/acii.2013.17]
74. Mori J, Prendinger H, Ishizuka M. Evaluation of an embodied conversational agent with affective behavior. In: Proceedings of the AAMAS03 Workshop on Embodied Conversational Characters as Individuals. 2003 Presented at: AAMAS03 Workshop on Embodied Conversational Characters as Individuals; 2003; Melbourne, Australia URL: http://www.miv.t.u-tokyo.ac.jp/papers/jmoriECA03.pdf

75. Johanson DL, Ahn HS, Sutherland CJ, Brown B, MacDonald BA, Lim JY, et al. Smiling and use of first-name by a healthcare receptionist robot: effects on user perceptions, attitudes, and behaviours. Paladyn J Behav Robot 2020;11(1):40-51 [FREE Full text] [doi: 10.1515/pjbr-2020-0008]
76. Chentsova-Dutton YE, Chu JP, Tsai JL, Rottenberg J, Gross JJ, Gotlib IH. Depression and emotional reactivity: variation among Asian Americans of East Asian descent and European Americans. J Abnorm Psychol 2007;116(4):776-785. [doi: 10.1037/0021-843x.116.4.776]
77. Tsai JL, Knutson B, Fung HH. Cultural variation in affect valuation. J Pers Soc Psychol 2006 Feb;90(2):288-307. [doi: 10.1037/0022-3514.90.2.288] [Medline: 16536652]
78. Ferstl Y, McDonnell R. A perceptual study on the manipulation of facial features for trait portrayal in virtual agents. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: IVA '18: International Conference on Intelligent Virtual Agents; November 5 - 8, 2018; Sydney NSW Australia p. 281-288. [doi: 10.1145/3267851.3267891]
79. Mori M. The uncanny valley. Energy 1970;7(4):33-35. [doi: 10.5749/j.ctvtv937f]
80. Johnson AM, DiDonato MD, Reisslein M. Animated agents in K-12 engineering outreach: preferred agent characteristics across age levels. Comput Hum Behav 2013 Jul;29(4):1807-1815. [doi: 10.1016/j.chb.2013.02.023]
81. Nass C, Isbister K, Lee E. Truth is beauty: researching embodied conversational agents. In: Cassell J, Sullivan J, Prevost S, Churchill E, editors. Embodied Conversational Agents. Cambridge, MA: MIT Press; 2000:374-402.
82. Benbasat I, Dimoka A, Pavlou P, Qiu L. Incorporating social presence in the design of the anthropomorphic interface of recommendation agents: insights from an fMRI study. In: Proceedings of the 31st International Conference on Information Systems (ICIS). 2010 Presented at: 31st International Conference on Information Systems (ICIS); 2010; St. Louis, Missouri URL: https://www.semanticscholar.org/paper/Incorporating-Social-Presence-in-the-Design-of-the-Benbasat-Dimoka/dab618d213b4c8cfe4447308d5c4538f859f13a6
83. Spence PR, Lachlan KA, Spates SA, Shelton AK, Lin X, Gentile CJ. Exploring the impact of ethnic identity through other-generated cues on perceptions of spokesperson credibility. Comput Hum Behav 2013;29(5):3-11. [doi: 10.1016/j.chb.2012.12.026]
84. Liao Y, He J. Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. In: Proceedings of the 25th International Conference on Intelligent User Interfaces. 2020 Presented at: IUI '20: 25th International Conference on Intelligent User Interfaces; March 17 - 20, 2020; Cagliari Italy p. 430-442. [doi: 10.1145/3377325.3377488]
85. Kaplan F. Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. Int J Hum Robot 2012;1(3):465-480. [doi: 10.1142/s0219843604000289]
86. Tamagawa R, Watson CI, Kuo IH, MacDonald BA, Broadbent E. The effects of synthesized voice accents on user perceptions of robots. Int J of Soc Robot 2011 Jun 2;3(3):253-262. [doi: 10.1007/s12369-011-0100-4]
87. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: trust-building factors and implications for conversational agent design. Int J Hum–Comput Interact 2020;37(1):81-96. [doi: 10.1080/10447318.2020.1807710]
88. Shamekhi A, Czerwinski M, Mark G, Novotny M, Bennett GA. An exploratory study toward the preferred conversational style for compatible virtual agents. In: Proceedings of the International Conference on Intelligent Virtual Agents. 2017 Presented at: International Conference on Intelligent Virtual Agents; August 27-30, 2017; Stockholm, Sweden p. 40-50. [doi: 10.1007/978-3-319-47665-0_4]
89. Thomas P, McDuff D, Czerwinski M, Craswell N. Expressions of style in information seeking conversation with an agent. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020 Presented at: SIGIR '20: The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval; July 25 - 30, 2020; Virtual Event China p. 1171-1180. [doi: 10.1145/3397271.3401127]
90. Hoegen R, Aneja D, McDuff D, Czerwinski M. An end-to-end conversational style matching agent. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019 Presented at: IVA '19: ACM International Conference on Intelligent Virtual Agents; July 2 - 5, 2019; Paris France p. 111-118. [doi: 10.1145/3308532.3329473]
91. Byrne D. Interpersonal attraction and attitude similarity. J Abnorm Soc Psychol 1961 May;62(3):713-715. [doi: 10.1037/h0044721] [Medline: 13875334]

Abbreviations

ANOVA: Analysis of Variance
CASA: Computers Are Social Actors
CONSORT: Consolidated Standards of Reporting Trials
RCIT: Relationship Closeness Induction Task

Edited by R Kukafka, G Eysenbach; submitted 21.05.21; peer-reviewed by V Ta, R Zhang; comments to author 30.07.21; revised version received 26.08.21; accepted 04.10.21; published 25.11.21.

Please cite as:
Loveys K, Sagar M, Zhang X, Fricchione G, Broadbent E
Effects of Emotional Expressiveness of a Female Digital Human on Loneliness, Stress, Perceived Support, and Closeness Across Genders: Randomized Controlled Trial
J Med Internet Res 2021;23(11):e30624
URL: https://www.jmir.org/2021/11/e30624
doi:10.2196/30624
PMID:34842540

Original Paper

Long-term Effectiveness of a Multistrategy Behavioral Intervention to Increase the Nutritional Quality of Primary School Students' Online Lunch Orders: 18-Month Follow-up of the Click & Crunch Cluster Randomized Controlled Trial

Rebecca Wyse[1,2,3,4], PhD; Tessa Delaney[1,2,3,4], BND; Fiona Stacey[1,2,3,4], PhD; Christophe Lecathelinais[1,2,3,4], DESS; Kylie Ball[5], PhD; Rachel Zoetemeyer[1,2], BND; Hannah Lamont[1,2], BND; Rachel Sutherland[1,2,3,4], PhD; Nicole Nathan[1,2,3,4], PhD; John H Wiggers[1,2,3,4], PhD; Luke Wolfenden[1,2,3,4], PhD

[1]Hunter New England Population Health, Wallsend, Australia

[2]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[3]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

[4]Hunter Medical Research Institute, New Lambton, Australia

[5]Institute for Physical Activity and Nutrition, Deakin University, Melbourne, Australia

Corresponding Author:
Rebecca Wyse, PhD
Hunter New England Population Health
Locked Bag 10
Wallsend, 2287
Australia
Phone: 61 2 404 20272
Email: r.wyse@newcastle.edu.au

Abstract

Background: School food services, including cafeterias and canteens, are an ideal setting in which to improve child nutrition. Online canteen ordering systems are increasingly common and provide unique opportunities to deliver choice architecture strategies to nudge users to select healthier items. Despite evidence of short-term effectiveness, there is little evidence regarding the long-term effectiveness of choice architecture interventions, particularly those delivered online.

Objective: This study determined the long-term effectiveness of a multistrategy behavioral intervention (Click & Crunch) embedded within an existing online school lunch-ordering system on the energy, saturated fat, sugar, and sodium content of primary school students' lunch orders 18 months after baseline.

Methods: This cluster randomized controlled trial (RCT) involved a cohort of 2207 students (aged 5-12 years) from 17 schools in New South Wales, Australia. Schools were randomized to receive either a multistrategy behavioral intervention or the control (usual online ordering only). The intervention strategies ran continuously for 14-16.5 months until the end of follow-up data collection. Trial primary outcomes (ie, mean total energy, saturated fat, sugar and sodium content of student online lunch orders) and secondary outcomes (ie, the proportion of online lunch order items that were categorized as *everyday*, *occasional*, and *caution*) were assessed over an 8-week period at baseline and 18-month follow-up.

Results: In all, 16 schools (94%) participated in the 18-month follow-up. Over time, from baseline to follow-up, relative to control orders, intervention orders had significantly lower energy (–74.1 kJ; 95% CI [–124.7, –23.4]; *P*=.006) and saturated fat (–0.4 g; 95% CI [–0.7, –0.1]; *P*=.003) but no significant differences in sugar or sodium content. Relative to control schools, the odds of purchasing *everyday* items increased significantly (odds ratio [OR] 1.2; 95% CI [1.1, 1.4]; *P*=.009, corresponding to a +3.8% change) and the odds of purchasing *caution* items significantly decreased among intervention schools (OR 0.7, 95% CI [0.6, 0.9]; *P*=.002, corresponding to a –2.6% change). There was no between-group difference over time in canteen revenue.

Conclusions: This is the first study to investigate the *sustained* effect of a choice architecture intervention delivered via an online canteen ordering systems in schools. The findings suggest that there are intervention effects up to 18-months postbaseline in terms of decreased energy and saturated fat content and changes in the relative proportions of healthy and unhealthy food purchased for student lunches. As such, this intervention approach may hold promise as a population health behavior change

strategy within schools and may have implications for the use of online food-ordering systems more generally; however, more research is required.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618000855224; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375075

(J Med Internet Res 2021;23(11):e31734) doi:10.2196/31734

KEYWORDS

child diet; consumer behavior; intervention; RCT; public health nutrition; obesity; school; school canteen; long-term follow-up; choice architecture; public health; nutrition; children; diet; eHealth; school lunch

Introduction

Poor diet is a leading cause of death and disability globally [1]. As dietary patterns in childhood track through to adulthood and are predictive of future disease [2], improving child nutrition is a global public health priority [3]. School food services, including cafeterias and canteens, are an ideal setting in which to improve child nutrition [4]. In Australia, school canteens are accessed by up to 95% of Australian children [5,6] to buy food and drinks during meal and snack times. Energy-dense, nutrient-poor foods and foods high in saturated fat, salt, or sugar are frequently available at school canteens and are commonly purchased by students [1,7-9]. As such, interventions to improve the nutritional quality of foods purchased at school canteens are warranted.

Choice architecture interventions that "nudge" people toward healthier behaviors by modifying the environment to increase the prominence or convenience of target food and drink items have shown promise in increasing the selection of healthy foods from school food service settings [10,11]. A recent systematic review of 29 choice architecture interventions within the school food service setting found that strategies such as point-of-purchase labeling, prompts, and food placement are positively associated with students' selection of healthier foods [11]. However, there remains little evidence of their longer-term impacts, with 21 of the 29 (72.4%) studies assessing dietary impacts over a period of less than 4 months [11]. Only two studies assessed the intervention impact 12 months or more postbaseline [12,13]. The cluster randomized controlled trial (RCT) by Larson et al [13] found increased participation in a school breakfast program relative to baseline, following the introduction of grab-and-go breakfast carts over a 2-year study, and the nonrandomized trial by Ensaff et al [12] found that an intervention involving repositioning, promoting, and labeling healthy target foods increased the selection of those items measured intermittently over 2 years.

Given that the effects of behavioral interventions typically attenuate over time [14,15], to achieve enduring improvements in public health nutrition, interventions need to be able to support long-term behavior change. Choice architecture interventions may be more resilient to attenuation over time, as they operate mainly through automatic psychological processes, and are not dependent on an ongoing cognitive load or self-regulatory skills of users [16]. However, little is known about their longer-term effects [17]. Although a recent long-term evaluation of the Healthy, Hunger-Free Kids Act of 2010 in US middle schools found that a multicomponent intervention, which included nudges in the form of defaults, found significant increases in the diet quality score, as measured by NHANES up to 4 years later [18], evidence is particularly scant regarding choice architecture interventions delivered online, with only 2 of the 29 (6.9%) studies in the above review delivered online [11] and both with a short follow-up period of only 2 weeks [19] and 2 months [20]. As such, more research is needed to determine the long-term effectiveness of online choice architecture interventions.

Online canteen ordering systems are common in Australian schools and allow parents and students (users) to order and pay for school lunches online. These systems provide unique opportunities to deliver choice architecture strategies to nudge users to select healthier items and assess their long-term impacts. The research team recently conducted the Click & Crunch trial to investigate the impact of choice architecture strategies embedded in an existing online lunch-ordering system in improving the nutritional quality of primary school students' online lunch order purchases [21]. At 12-month follow-up, the intervention significantly lowered the energy (–69 kJ) and saturated fat (–0.6 g) content of student lunch orders (P=.01) without any adverse impact on canteen revenue [21]. Additionally, a higher proportion of healthy or *everyday* items (odds ratio [OR] 1.69; P<.001, +9.8%) and a lower proportion of less healthy or *occasional* items (OR 0.68; P<.001, –7.7%) were purchased by students at intervention schools compared with controls [21]. While these initial outcomes are promising, an assessment of the longer-term impact of the intervention on student lunch purchases is needed to better quantify its contribution to public health nutrition.

The primary aim of this study was to determine the long-term effectiveness from baseline to 18-month follow-up of the Click & Crunch intervention, a multistrategy behavioral intervention embedded within an existing online lunch-ordering system, in reducing the energy, saturated fat, sugar, and sodium content of primary school students' lunch orders.

Methods

Subjects and Methods

A detailed description of the trial methods is provided in the study protocol [22]. The study was approved and procedures monitored by the Human Research Ethics Committee of the University of Newcastle (reference no. H-2017-0402) and the relevant New South Wales (NSW) Catholic Schools Dioceses (including Sydney, Parramatta, Lismore, Maitland-Newcastle, Bathurst, Canberra-Goulburn, Wagga Wagga, Wollongong, and

Wilcannia-Forbes). The trial is reported according to the Consolidated Standards of Reporting Trials (CONSORT) guidelines for cluster RCTs (see Multimedia Appendix 1). The original trial methods and 12-month follow-up were prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12618000855224). The 18-month follow-up was conducted in accordance with previously registered procedures, and all outcomes were registered.

Design

The trial used a parallel cluster RCT design. Schools with an existing online lunch-ordering system were randomly assigned either to a multistrategy behavioral intervention or a control (standard online ordering system). Outcome data were collected over an 8-week period (weeks 1-8 of the school term) at baseline and again 12 months and 18 months after study initiation. This manuscript reports the 18-month findings. Once the intervention commenced and was switched on within the online ordering system, the intervention strategies remained in place until after collection of the 18-month follow-up data.

Sample

Schools

A sample of 17 (9 intervention, 8 control) nongovernment (Catholic and Independent) schools catering for students aged 5-12 years in NSW, Australia, that were existing users of the Flexischools online school lunch-ordering system were recruited. Schools were ineligible if they (1) had used the system for less than 1 month prior to recruitment, (2) were privately operated (as these externally leased canteens often service multiple schools, increasing the risk of intervention contamination), or (3) catered exclusively for secondary students (due to differences in the NSW Healthy School Canteen Strategy). Combined schools that catered for both primary and secondary students were only eligible to participate if they had a separate menu for primary students (as the NSW Healthy School Canteen Strategy varies by these age groups). Schools were approached to participate by mail and telephone.

Users

All students from kindergarten to grade 5 who placed an online lunch order during baseline (term 2, May-July 2018) were included. Students were ineligible if they were in grade 6 or a 5/6 composite class at baseline, as they would have left the school prior to the collection of follow-up data, as were all other nonstudent users of the online ordering system, such as teaching staff and guests.

Orders

Only orders placed on a mobile device were included, as users placing orders via a desktop device were not exposed to all intervention strategies. Recurring orders placed prior to intervention commencement were excluded, as were orders placed for food service periods other than lunch (eg, recess or special food days), as users would not have been exposed to all intervention strategies. Given that the system offers the capability for users to purchase items in bulk (eg, for class parties), orders that contained 15 or more items were also excluded. This decision was based on dietitian consensus and knowledge of the number of items that are typically in online lunch orders (eg, on average, a primary school student lunch order contains 2.3 items). Small items, such as chicken nuggets, may be sold individually and then purchased in higher quantities (eg, 6 or 12 nuggets), accounting for the high upper limit for a plausible order.

Randomization and Blinding

After recruitment and following the collection of the baseline canteen menus, an independent statistician block-randomized schools (block size ranging from 2 to 4) using a random number function in Microsoft Excel to an intervention or a control group in a 1:1 ratio. Randomization was stratified by school sector (Catholic vs Independent) and socioeconomic status (most vs least advantaged) based on the school postcode (Socio-economic Indexes for Areas [SEIFA]) [23].

Intervention

The intervention is described in full elsewhere [22]. The intervention was guided by principles of choice architecture [16] and used strategies that have been demonstrated to support healthier food choices in similar food service settings [24-29]. Briefly, intervention schools had a series of choice architecture strategies applied to their online menu within the online canteen ordering system. The point at which each intervention school had the strategies switched on varied from August to October 2018, but once switched on, they remained in place for 14-16.5 months until the end of the 18-month follow-up period. The intervention sought to encourage the purchase of healthier foods and beverages aligned to the Australian Dietary Guidelines and NSW Healthy School Canteen Strategy, that is, the intervention encouraged the ordering of food items lower in energy, saturated fat, sugar, or sodium. The intervention strategies included:

- *Menu labeling:* Each menu item was labeled with colored symbols as *everyday*, *occasional*, or *caution* (also known as "should not be sold") per the NSW Healthy School Canteen Strategy [30], and a key defining the symbols was provided.
- *Positioning: Everyday* menu items and healthier categories (eg, salads and fruit) were positioned most prominently (ie, first), with *caution* and *occasional* items positioned least prominently (ie, middle and last, respectively). *Occasional* and *caution* items with multiple flavors (eg, chips) were displayed on a separate screen that users had to click through to reach.
- *Prompting:* Healthier categories received an attractive image and a text prompt (ie, "This is a good choice."). Users selecting *occasional* or *caution* hot foods received a prompt to also select fruit, vegetables, or water ("Healthy add-ons").
- *Feedback:* Prior to users finalizing their orders, they were provided with tailored feedback (ie, a pie graph and accompanying text) based on the proportion of *everyday* items in their order.
- *Incentives:* A reward symbol cartoon character and congratulatory text were printed on the student's lunch order bag for orders that contained 100% *everyday* items.

Canteen Supportive Strategy

In addition, an audit feedback report was emailed to canteen managers and principals at the start of the intervention period, classifying each menu item as *everyday*, *occasional*, or *caution* and providing feedback on substitutable healthier items. It also provided general information about how to price items to encourage healthier purchasing. This targeted the canteen managers rather than parent and student users.

With the exception of the feedback report for canteen managers, all strategies were incorporated directly into the school's online menu within the online canteen ordering system.

Intervention Fidelity

Once per term during the intervention period (ie, approximately every 10 weeks), a research assistant (author RZ) monitored the online menus to record adherence to the intervention strategies. Where unlabeled menu items (eg, new items) were identified, the research team contacted the online lunch-ordering provider to apply the label and intervention strategies accordingly. Given that three of the four other online strategies were programmed based on the label that was applied, correct application of the label ensured these strategies were also correctly implemented.

Control

The control group did not receive any of the intervention strategies and had access to the standard online ordering system only.

Data Collection and Outcomes

Purchasing data were automatically collected and stored by the online ordering system and were subsequently extracted for the defined baseline and follow-up data collection periods by Flexischools. The purchasing data for the baseline period were retrospectively collected. Data were collected over three 8-week periods spanning an 18-month period, with baseline occurring during May to July 2018 and the 18-month follow-up occurring during October to December 2019 (baseline: term 2, 2018; 12 months: term 2, 2019; and 18 months: term 4, 2019). The 12-month follow-up was the primary trial endpoint, and 12-month outcomes have been previously published [21]. This paper reports the baseline and 18-month follow-up only in order to examine intervention effectiveness in the longer term.

Primary Trial Outcomes

The primary trial outcomes at 18 months were identical to those at 12 months and included the mean total energy (kJ), saturated fat (g), sugar (g), and sodium (mg) content of online lunch orders. These nutrition outcomes were calculated by a dietitian who conducted a comprehensive menu assessment of each school's canteen menu and applied those values to the online purchasing data that were automatically collected and stored by the Flexischools online canteen ordering system. The dietitian conducted the menu assessment by a telephone interview with each canteen manager and collected the brand, product name, serving size, or recipe for each available item. After the interview, the dietitian generated a nutritional profile for each item. For canteen-made products, the dietitian entered the recipe into FoodWorks nutrition analysis version 9 (Xyris Software) [31]. To assign the nutritional profile (energy, saturated fat, sugar, and sodium content) for prepackaged items, the dietitian consulted a series of sources in the following order: (1) a database of over 2000 commonly stocked canteen products developed by the researchers over the past decade, (2) the FoodFinder database [32], (3) the FoodSwitch website [33], and (4) an internet search for the product's nutrient panel.

Secondary Trial Outcomes

The secondary trial outcomes were as follows:

- *Healthy purchasing outcomes:* The proportion of all online lunch order items that were (1) *everyday*, (2) *occasional*, and (3) *caution*, as classified by the dietitian using the NSW Healthy School Canteen Strategy [30].
- *Revenue (adverse outcome):* Automatically collected purchasing data were used to determine impact of the intervention on the school canteen's average weekly revenue in Australian dollars.

School Characteristics

At baseline, descriptive data regarding school characteristics, such as the number of student enrolments, the proportion of Aboriginal and Torres Strait Islander enrolments, and postcode were collected from a national website (myschool.edu.au) [34].

Canteen Characteristics

To describe the sample, data regarding canteen operations (eg, number of days open, model of operation, paid canteen manager) were collected during a canteen manager survey during the initial 12-month follow-up.

Sample Size Calculation

Sample size estimates were calculated a priori based on a 12-month follow-up and are also described in the trial protocol [22]. The original sample size indicated that a sample of 26 schools was required to ensure a detectable difference of 195 kJ per lunch order, with 80% power, an intraclass correlation coefficient (ICC) of 0.05, and a 0.0125 significance level at 12-month follow-up (Bonferroni-adjusted).

Statistical Analysis

All outcome data were analyzed under an intention-to-treat approach whereby all student lunch orders and schools were analyzed based on the groups they were originally allocated, and included data from students that had baseline purchasing data.

Primary trial outcomes were assessed using separate linear mixed models by comparing differences between intervention and control groups over time (from baseline to 18 months) through the inclusion of a group-by-time interaction fixed effect. All models included a random intercept for schools (to account for potential school-level clustering), a nested random intercept and random time effect for students (to account for repeat measurements within and over time), and fixed effects for the school sector and SEIFA. All available data were incorporated into the model (baseline, 12 months, and 18 months). The results of the 12-month follow-up are included as a supplementary file (see Multimedia Appendix 2). The unit of analysis for primary

trial outcomes was lunch orders, where a lunch order could contain multiple items.

Secondary trial outcomes relating to nutritional quality (*everyday*, *occasional*, *caution*) were assessed using separate logistic mixed models. Changes in the proportion of items ordered belonging to each category (ie, *everyday* items vs *not everyday* items) were compared between intervention and control groups over time by including a group-by-time interaction fixed effect. Similar to primary trial outcomes, all models included a random intercept for schools (to account for potential school-level clustering), a nested random intercept and random time effect for students (to account for repeat measurements), and fixed effects for the school sector and SEIFA and included the 12-month outcomes. Differences in the average weekly revenue were assessed using a linear mixed model with a similar structure to models for the primary outcomes. School and canteen characteristics were previously reported in the 12-month outcome paper and are included here for context.

A per-protocol analysis was also conducted, which included only those schools that had at least 80% of verifiable strategies correctly applied at follow-up. Statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC, USA). As no differences in subgroups (with respect to student grade, school sector, or frequency of canteen use) were found in the 12-month follow-up [21], no analyses were conducted on 18-month data.

Results

Sample Characteristics

The baseline characteristics of the sample are presented in Tables 1 and 2. Characteristics were similar between groups; however, intervention schools had higher student enrolments and more lunch orders per week than control schools (no significance testing). In total, 1042 intervention students (8 schools) and 667 control students (8 schools) from a total of 16 schools were included in the 18-month follow-up, representing 94% of the 17 schools recruited to the trial (see Figure 1). One intervention school withdrew after the 12-month follow-up, but prior to the 18-month follow-up, and contributed data for two of the three time points in the analysis. The sample also contained two schools with privately operated (externally leased) canteens that did not initially identify as such in the recruitment process. The schools were retained as there was no contamination risk, given that they did not service any other schools in the sample. One was allocated to the control group and one to the intervention group. There were 1435 recurring orders (4.5%) that did not meet the eligibility criteria and were removed from analysis, which resulted in 10 students who only had recurring orders being excluded. Furthermore, four orders were excluded due to being implausibly large based on dietitian assessment.

Table 1. School characteristics of the sample at baseline for all participating schools by group [21].

School characteristics reported at baseline	All schools (n=17)	Intervention schools (n=9)	Control schools (n=8)
School sector, n (%)			
Independent	7 (41.2%)	4 (44.4%)	3 (37.5%)
Catholic	10 (58.8%)	5 (55.6%)	5 (62.5%)
Mean (SD) number of enrolments[a]	443.7 (177.4)	501.3 (207.9)	386 (134.3)
Mean % of Aboriginal or Torres Strait Islander students	5%	6%	4%
Socioeconomic status, n (%)			
Least advantaged	7 (41.2%)	3 (33.3%)	4 (50%)
Most advantaged	10 (58.8%)	6 (66.7%)	4 (50%)

[a]Excluding combined schools (as this information was not available on the MySchool website)

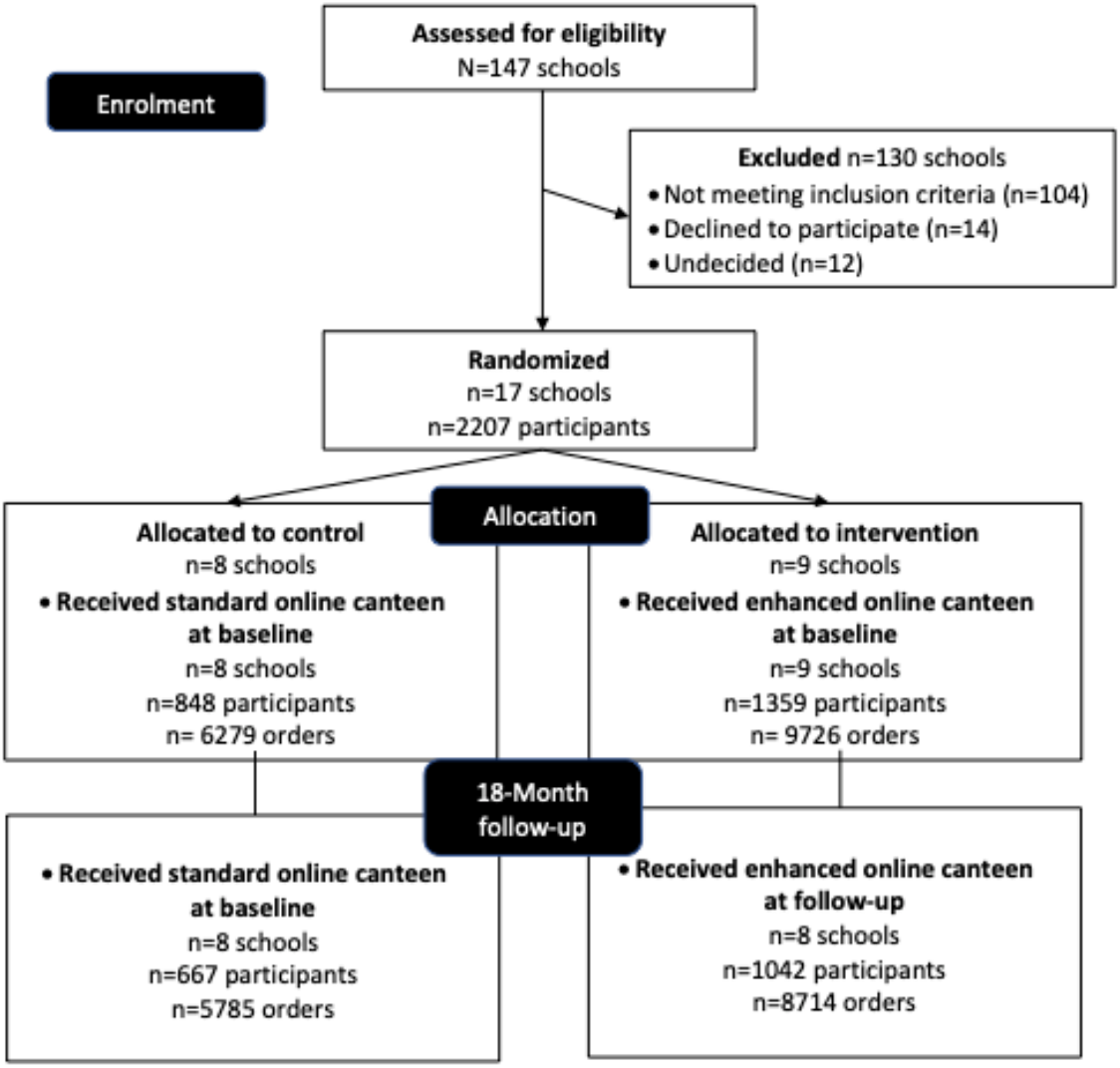
Table 2. Canteen characteristics of the sample at baseline for all participating schools by group [21].

Characteristics reported at baseline	All schools (n=15)	Intervention schools (n=7)	Control schools (n=8)
Type of operation, n (%)			
Principal/school run	12 (80%)	6 (85.7%)	6 (75%)
P&F[a]/P&C[b] association run	1 (6.7%)	0 (0%)	1 (12.5%)
Contracted food service	2 (13.3%)	1 (14.3%)	1 (12.5%)
Type of manager, n (%)			
Paid	15 (100%)	7 (100%)	8 (100%)
Volunteer	0 (0%)	0 (0%)	0 (0%)
Days of operation, n (%)			
5 days a week	11 (78.6%)	6 (85.7%)	5 (62.5%)
3-4 days a week	3 (20%)	1 (14.3%)	2 (25%)
1-2 days a week	1 (6.7%	0 (0%)	1 (12.5%)
Mean (SD) number of weekly online lunch orders (per school)	—	135.9 (80.3)	98.3 (91.3)

[a]P&F: Parents and Friends.

[b]P&C: Parents and Citizens.

Figure 1. CONSORT diagram. CONSORT: Consolidated Standards of Reporting Trials.



Outcomes

In the intervention group, from baseline to 18-month follow-up, online lunch orders contained significantly less energy (–74.1 kJ; 95% CI [–124.7, –23.4]; *P*=.006) and saturated fat (–0.4 g; 95% CI [–0.7, –0.1]; *P*=.003) relative to control orders (see Tables 3-5). There was also a significant between-group difference over time in the odds of lunch orders including *everyday* items (OR 1.2; 95% CI [1.1, 1.4]; *P*=.009) corresponding to a 3.8% increase in *everyday* items purchased among the intervention group, and a significant difference in the odds of lunch orders including *caution* items (OR 0.7; 95% CI [0.6, 0.9]; *P*=.002) corresponding to a 2.6% decrease in *caution* items among the intervention group. There was no between-group difference over time in the sugar (–0.5 g; 95% CI [–1.7, 0.7]; *P*=.4) or sodium (–3.0 mg; 95% CI [–28.0, 22.1]; *P*=.8) content of lunch orders or in the odds of lunch orders including *occasional* items (OR 1.0; 95% CI [0.8, 1.1]; *P*=.6).

Table 3. Primary outcomes from baseline to 18-month follow-up.

Primary outcomes	Baseline intervention (n=23,526 items; n=9726 orders; n=1359 children; n=9 schools)	Baseline control (n=14,124 items; n=6279 orders; n=848 children; n=8 schools)	18-month follow-up intervention (n=20,351 items; n=8714 orders; n=1042 children; n=8 schools)	18-month follow-up control (n=12,579 items; n=5785 orders; n=667 children; n=8 schools)
Energy (kJ) mean (SD)	1634.4 (704.2)	1632.1 (743.0)	1603.8 (700.4)	1671.4 (876.1)
Saturated fat (g) mean (SD)	5.2 (3.9)	4.6 (3.2)	4.6 (3.7)	4.3 (3.3)
Sugar (g) mean (SD)	12.9 (14.0)	15.8 (19.1)	13.1 (13.9)	17.3 (24.4)
Sodium (mg) mean (SD)	596.1 (343.0)	599.3 (328.9)	606.1 (409.1)	590.3 (344.3)

Table 4. Secondary outcomes from baseline to 18-month follow-up.

Secondary outcomes	Baseline intervention (n=23,526 items; n=9726 orders; n=1359 children; n=9 schools)	Baseline control (n=14,124 items; n=6279 orders; n=848 children; n=8 schools)	18-month follow-up intervention (n=20,351 items; n=8714 orders; n=1042 children; n=8 schools)	18-month follow-up control (n=12,579 items; n=5785 orders; n=667 children; n=8 schools)
% of student lunch order items classified as *everyday*, % (n)	31.6% (7423)	40.4% (5711)	41.5% (8439)	46.6% (5859)
% of student lunch order items classified as *occasional*, % (n)	47.9% (11261)	43.8% (6185)	43.5% (8846)	40.2% (5052)
% of student lunch order items classified as *caution*, % (n)	20.6% (4842)	15.8% (2228)	15.1% (3066)	13.3% (1668)
% of energy from saturated fat, mean (SD)	11.0 (5.9)	9.9 (5.1)	9.8 (6.1)	9.3 (5.1)
% of energy from sugar, mean (SD)	12.0 (11.8)	13.9 (12.7)	12.8 (13.0)	15.7 (15.0)

The per-protocol analysis excluded three of the eight intervention schools that only partially applied the intervention strategies. The pattern of results for the primary outcomes was similar between the per-protocol and main analyses (see Table 5), with significant differences observed for the energy and saturated fat content of lunch orders and slightly larger effect sizes (–93 kJ vs –74.1 kJ). There were no significant differences in the secondary outcomes in the per-protocol analysis at the prespecified Bonferroni-adjusted significance level (*P*=.0125).

Table 5. Main vs per-protocol analysis from baseline to 18-month follow-up[a].

Outcomes	Main analysis differential effect (group by time) (95% CI)	Main analysis OR[b] (95% CI)	Main analysis *P* value	Per-protocol analysis differential effect (group by time) (95% CI)	Per-protocol analysis OR (95% CI)	Per-protocol analysis *P* value
Primary outcomes						
Energy (kJ)	–74.1 (–124.7, –23.4)	—	0.006	–93.0 (–151.9, –34.2)	—	0.003
Saturated fat (g)	–0.4 (–0.7, –0.1)	—	0.003	–0.5 (–0.8, –0.2)	—	0.003
Sugar (g)	–0.5 (–1.7, 0.7)	—	0.39	–0.1 (–1.5, 1.3)	—	0.87
Sodium (mg)	–3.0 (–28.0, 22.1)	—	0.81	7.3 (–21.9, 36.4)	—	0.61
Secondary outcomes						
% of student lunch order items classified as *everyday*, % (n)	3.8%	1.2 (1.1, 1.4)	0.009	0.6%	1.0 (0.9, 1.1)	0.68
% of student lunch order items classified as *occasional*, % (n)	–1.1%	1.0 (0.8, 1.1)	0.64	2.6%	1.1 (1.0, 1.3)	0.013
% of student lunch order items classified as *caution*, % (n)	–2.6%	0.7 (0.6, 0.9)	0.002	–2.6%	0.9 (0.8, 1.0)	0.07
% of energy from saturated fat	–0.4 (–0.8, –0.02)	—	0.039	–0.6 (–1.1, –0.1)	—	0.02
% of energy from sugar	–0.08 (–1.0, 0.8)	—	0.86	0.8 (–0.3, 1.8)	—	0.14

[a]All models included a random intercept for school, a nested random intercept and random time effect for students, and fixed effects for sector and SEIFA. All available data was incorporated into the model (baseline, 12-months, 18-months) to describe purchasing patterns over time.

[b]OR: odds ratio.

Revenue

Analysis of the average weekly canteen revenue indicated that over time from baseline to 18-month follow-up, there were no significant differences between the intervention and control groups ($80.42; 95% CI [–104.48, 265.33]; *P*=.4). This finding was unchanged in the per-protocol analysis (see Table 6).

Table 6. Average weekly revenue per school ($[a]).

Baseline intervention, mean (SD)	Baseline control, mean (SD)	18-month follow-up intervention, mean (SD)	18-month follow-up control, mean (SD)	Main analysis differential effect (group by time) (95% CI)	Main analysis *P* value	Per-protocol analysis differential effect (group by time) (95% CI)	Per-protocol analysis *P* value
$668.61 ($420.90)	$496.10 ($442.63)	$1081.03 ($525.54)	$758.76 ($576.13)	$80.42 (–104.48, 265.33)	0.39	$154.56 (–59.15, 368.26)	0.16

[a]All $ amounts are in AUD $. A currency exchange rate of AUD $1=US $0.75 is applicable.

Quality and Fidelity Checks

Quality checks that were conducted immediately after switching on the intervention identified a technical glitch whereby purchasing a menu item routinely sold in multiples (eg, 6 × chicken nuggets) affected the application of the add-on strategy. The solution was to turn off the strategy for items routinely sold in multiples. Routine fidelity checks conducted once per term during the intervention period did not identify any further technical errors. The fidelity checks indicated that the menu labels were correctly applied to 93%-95% of all online menu items in the first 4 terms of the trial (12 months [21]) and 97% of all items during the last 2 terms (12-18 months). The strategies of positioning, tailored feedback, and incentives were automatically programmed based on the label assigned to each item, and as such, the fidelity checks for the menu labels also applied for these strategies. As previously reported [21], there was a programming issue with the healthy add-on strategies for items sold in multiples (eg, chicken nuggets), and as a result, this strategy was removed only from these items, and two intervention schools in this follow-up had previously requested this strategy be switched off entirely. The presence of the incentive strategy was verified in five of the eight participating intervention schools.

Availability

The proportion of *everyday*, *occasional*, and *caution* menu items was similar between Independent and Catholic school menus at baseline (59% and 59% *everyday* items, 15% and 16% *occasional* items, and 26% and 25% *caution* items, respectively). Furthermore, the proportion of *everyday*, *occasional*, and *caution* menu items was similar between intervention and

control menus at baseline (58% and 61% *everyday* items, 16% and 15% *occasional* items, and 22% and 25% *caution* items, respectively).

Discussion

Principal Results

This long-term follow-up of the Click & Crunch intervention using automatically collected purchasing data found significant between-group differences over time from baseline to 18-month follow-up. Specifically, the energy and saturated fat content of intervention lunch orders was significantly lower than controls (–74.1 kJ of energy; –0.4 g of saturated fat). There were no significant between-group differences with respect to sodium or sugar content or the percentage of energy from sugar or fat. Among intervention schools relative to control schools, from baseline to 18-month follow-up, the odds of orders containing *everyday* items were significantly higher (OR 1.2, *P*=.009, corresponding to a 3.8% increase in the purchase of these items), and the odds of orders containing *caution* items were significantly lower (OR 0.7, *P*=.002, corresponding to a 2.6% decrease in the purchase of these items). As such, the results suggest that at 18-month follow-up, the Click & Crunch intervention is effective in reducing the energy and saturated fat content of students' online lunch orders and in reducing the proportion of unhealthy items and increasing the proportion of healthy items purchased.

Comparison With Prior Work

The pattern of results was similar between the 12-month and 18-month follow-up. At 12 months (previously reported [21]), the Click & Crunch intervention was effective in reducing the energy and saturated fat content of student lunch orders, and the effects were similar in magnitude to those observed at 18-month follow-up (energy: –69 kJ at 12 months [21] and –74.1 kJ at 18 months; saturated fat: –0.6 g at 12 months [21] and –0.4 g at 18 months). Similarly, there were significant increases in the proportion of *everyday* items purchased among the intervention group (+9.8% at 12 months; +3.8% at 18 months) and significant decreases in the proportion of less healthy items purchased (–7.7% *occasional* items at 12 months; –2.6% *caution* items at 18 months).

There is limited research examining the impact of similar interventions to enable direct comparison of the long-term effects. Multiple systematic reviews of choice architecture interventions to improve dietary outcomes have highlighted the lack of research into long-term intervention outcomes [35,36], including the review of nudge interventions in schools, which reported there are significant knowledge gaps regarding the long-term impact of nudges within the school food environment [12]. However, a nonrandomized longitudinal study investigated the effects of choice architecture strategies (menu labeling and item repositioning) within the physical environment of a hospital cafeteria and recorded sales data from adult employees over a 2-year period [24]. After 2 years, the proportion of healthy items purchased increased by 5% and the proportion of unhealthy items purchased decreased by 4% (*P*<.001) [24], finding changes similar in magnitude to the current study and not providing evidence of label-fatigue over time. Both the school-based studies by Ensaff et al [12] and Larson et al [13] found improvement in the primary outcomes: section of target healthy items and participation in a breakfast program, respectively. Despite study limitations, collectively these findings provide important insights into the durability of choice architecture interventions, suggesting that unlike interventions that require controlled processing, such interventions may be sustained in the long term. As such, they may represent an attractive option for those interested in achieving long-term improvements in public health nutrition via school-based interventions.

Importantly, the magnitude of the effects for energy content and the proportion of *everyday* and *caution* items seen at 18 months appear to have public health significance. The study by Thorndike et al [24] modeled the observed changes in purchasing patterns in a hospital cafeteria 2 years after a choice architecture intervention was implemented and found a 6.2% decrease in total calories purchased and a 4% increase in calories from healthy food purchases [24]. Modeled data on high-frequency cafeteria users suggested that the long-term effects of this intervention could have an impact on obesity rates [24]. However, the change in kilojoules purchased reported in this study and used as the basis for modeling was larger than in the current study. A modeling study based on data from Australian children suggested that a decrease of 100 kcal/day (~418 kJ) would be sufficient to halve the current prevalence of overweight/obesity within a short period (less than 2 years) [37]. As such, the observed long-term decrease in lunch order kilojoule content (–74.1 kJ) is insufficient in isolation to decrease overweight/obesity rates. However, given the relative simplicity of the intervention and the potential for a wide reach, this intervention may play an important role as part of a suite of interventions adopted across multiple settings (school, after-school care and childcare, home, sporting and community clubs, etc) as are commonly adopted by governments in Australia and internationally [38].

Limitations and Strengths

The limitations of this study are similar to those described in the 12-month outcomes paper [21] and include the relatively small number of schools, the use of purchasing data rather than consumption data, the lack of individual demographic data, and the exclusion of government schools. It also should be acknowledged that the impact of a menu-labeling system on nutritional outcomes is dependent, in part, on the alignment of the labeling system with the target.

Study strengths include a rigorous cluster randomized controlled design, an excellent school retention rate (94%), and a large number of student participants. Furthermore, the evaluation was based on long-term, objective, and real-world purchasing data, which were independently verified, and menu assessments were based on gold-standard processes [39]. Data were also collected at multiple time points, indicating high intervention fidelity.

Conclusions

To the best of our knowledge, this is the first RCT to investigate the long-term effectiveness of choice architecture strategies applied online. The findings are encouraging and suggest that there are enduring intervention effects up to 18-months

postbaseline, including a difference of –74.1 kJ in the energy content and a difference of +3.8% in the proportion of healthy *everyday* items purchased. This provides much needed evidence about the sustainability of multistrategy choice architecture interventions (including menu labeling, positioning, prompting, feedback, and incentives) on children's school lunch ordering from online canteens. Although the intervention only produced modest effects, given the wide reach of online canteen ordering systems, it may be useful as one of a range of interventions to supplement existing strategies used to improve child diet within the school setting. Further research is required to determine whether the effects transfer to related online food-ordering settings, including groceries, fast food, and meal subscriptions, which are currently used by more than 1.2 billion people worldwide.

Acknowledgments

We would like to thank Flexischools, the Research Advisory Group, and the participating schools, canteen managers, and canteen users. The food composition data available on the Healthy Food Finder were provided by SP Health Co. Pty Ltd, trading as Digital Wellness (ABN 73 123 248 046) in collaboration with the New South Wales (NSW) Ministry of Health. The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request, pending ethics approval.

Authors' Contributions

RW conceived of the study; RW, LW, KB, JHW, and TD developed the overall research plan; RW, LW, NN, and RS provided study oversight; RW, TD, RZ, HL, LW, and FS conducted research; CL, FS, and TD analyzed data; CL provided statistical advice; and RW led the writing of the paper, with all authors contributing to paper revisions. RW has primary responsibility for the final content.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report. This work is supported by the National Health and Medical Research Council (NHMRC; grant no. APP1120233). In-kind support was provided by Hunter New England Population Health and the Hunter Medical Research Institute. RW is supported by a Heart Foundation postdoctoral fellowship (ID: 102156). LW receives salary support from an NHMRC career development fellowship (ID: APP1128348) and a Heart Foundation future leader fellowship (ID: 101175). Neither the NHMRC nor the Heart Foundation had any role in the design of the study, data collection, analysis or interpretation, or dissemination of findings. The provider (Flexischools) was selected through a competitive tender process. Flexischools is a commercial organization that provided the online canteen ordering infrastructure to schools that were included in the study. Flexischools had no role in the study design, data analysis, data interpretation, or writing of the manuscript. No financial disclosures were reported by the authors of this paper.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1174 KB - jmir_v23i11e31734_app1.pdf]

Multimedia Appendix 2
Results from the 12-month follow-up.
[DOCX File , 22 KB - jmir_v23i11e31734_app2.docx]

References

1. GBD 2017 Risk Factor Collaborator. Global, regional, and national comparative risk assessment of 84 behavioral, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1923-1994. [doi: 10.1016/S0140-6736(18)32225-6]
2. Maynard M, Gunnell D, Emmett P, Frankel S, Davey Smith G. Fruit, vegetables, and antioxidants in childhood and risk of adult cancer: the Boyd Orr cohort. J Epidemiol Community Health 2003;57(3):218-225 [FREE Full text] [doi: 10.1136/jech.57.3.218] [Medline: 12594199]
3. World Health Organization. Global Strategy on Diet, Physical Activity and Health. 2004. URL: https://www.who.int/publications/i/item/9241592222 [accessed 2020-11-16]
4. World Health Organization. Population-Based Approaches to Childhood Obesity Prevention. 2012. URL: https://apps.who.int/iris/handle/10665/80149 [accessed 2020-11-16]
5. Bell AC, Swinburn BA. What are the key food groups to target for preventing obesity and improving nutrition in schools? Eur J Clin Nutr 2004 Mar 29;58(2):258-263. [doi: 10.1038/sj.ejcn.1601775] [Medline: 14749745]
6. Hardy L. NSW Schools Physical Activity and Nutrition Survey (SPANS) 2015: Full Report. Sydney: NSW Department of Health; 2016.

7. Delaney T, Sutherland R, Wyse R, Wolfenden L, Lecathelinais C, Janssen L, et al. A cross-sectional study of the nutritional quality of student canteen purchases from New South Wales primary-school canteens. Public Health Nutr 2019 Jul 25;22(16):3092-3100. [doi: 10.1017/s1368980019001903]
8. Finch M, Sutherland R, Harrison M, Collins C. Canteen purchasing practices of year 1-6 primary school children and association with SES and weight status. Aust N Z J Public Health 2006;30(3):247-251.
9. Clinton-McHarg T, Janssen L, Delaney T, Reilly K, Regan T, Nathan N, et al. Availability of food and beverage items on school canteen menus and association with items purchased by children of primary-school age. Public Health Nutr 2018 Aug 02;21(15):2907-2914. [doi: 10.1017/s1368980018001726]
10. Marcano-Olivier MI, Horne PJ, Viktor S, Erjavec M. Using nudges to promote healthy food choices in the school dining room: a systematic review of previous investigations. J Sch Health 2020 Feb 18;90(2):143-157. [doi: 10.1111/josh.12861] [Medline: 31852016]
11. Metcalfe JJ, Ellison B, Hamdi N, Richardson R, Prescott MP. A systematic review of school meal nudge interventions to improve youth food behaviors. Int J Behav Nutr Phys Act 2020 Jun 19;17(1):77 [FREE Full text] [doi: 10.1186/s12966-020-00983-y] [Medline: 32560731]
12. Ensaff H, Homer M, Sahota P, Braybrook D, Coan S, McLeod H. Food choice architecture: an intervention in a secondary school and its impact on students' plant-based food choices. Nutrients 2015 Jun 02;7(6):4426-4437 [FREE Full text] [doi: 10.3390/nu7064426] [Medline: 26043039]
13. Larson N, Wang Q, Grannon K, Wei S, Nanney MS, Caspi C. A low-cost, grab-and-go breakfast intervention for rural high school students: changes in school breakfast program participation among at-risk students in Minnesota. J Nutr Educ Behav 2018 Feb;50(2):125-132.e1 [FREE Full text] [doi: 10.1016/j.jneb.2017.08.001] [Medline: 28951057]
14. Dave J. Assessing long-term impact of dietary interventions: important but challenging. Am J Clin Nutr 2019 Jan 01;109(1):5-6. [doi: 10.1093/ajcn/nqy339] [Medline: 30624574]
15. Yavuz HM, van Ijzendoorn MH, Mesman J, van der Veek S. Interventions aimed at reducing obesity in early childhood: a meta-analysis of programs that involve parents. J Child Psychol Psychiatry 2015 Jun 08;56(6):677-692. [doi: 10.1111/jcpp.12330] [Medline: 25292319]
16. Hollands GJ, Shemilt I, Marteau TM, Jebb SA, Kelly MP, Nakamura R, et al. Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. BMC Public Health 2013 Dec 21;13:1218 [FREE Full text] [doi: 10.1186/1471-2458-13-1218] [Medline: 24359583]
17. Marchiori DR, Adriaanse MA, De Ridder DT. Unresolved questions in nudging research: putting the psychology back in nudging. Soc Personal Psychol Compass 2017 Jan 13;11(1):e12297. [doi: 10.1111/spc3.12297]
18. Kinderknecht K, Harris C, Jones-Smith J. Association of the Healthy, Hunger-Free Kids Act with dietary quality among children in the US National School Lunch Program. JAMA 2020 Jul 28;324(4):359-368 [FREE Full text] [doi: 10.1001/jama.2020.9517] [Medline: 32721008]
19. Miller GF, Gupta S, Kropp JD, Grogan KA, Mathews A. The effects of pre-ordering and behavioral nudges on National School Lunch Program participants' food item selection. J Econ Psychol 2016 Aug;55:4-16. [doi: 10.1016/j.joep.2016.02.010]
20. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, et al. Cluster randomized controlled trial of a consumer behavior intervention to improve healthy food purchases from online canteens. Am J Clin Nutr 2017 Nov 27;106(5):1311-1320. [doi: 10.3945/ajcn.117.158329] [Medline: 28971849]
21. Wyse R, Delaney T, Stacey F, Zoetemeyer R, Lecathelinais C, Lamont H, et al. Effectiveness of a Multistrategy Behavioral Intervention to Increase the Nutritional Quality of Primary School Students' Web-Based Canteen Lunch Orders (Click & Crunch): Cluster Randomized Controlled Trial. J Med Internet Res 2021 Sep 07;23(9):e26054 [FREE Full text] [doi: 10.2196/26054] [Medline: 34491207]
22. Wyse R, Delaney T, Gibbins P, Ball K, Campbell K, Yoong SL, et al. Cluster randomised controlled trial of an online intervention to improve healthy food purchases from primary school canteens: a study protocol of the 'click & crunch' trial. BMJ Open 2019 Sep 05;9(9):e030538. [doi: 10.1136/bmjopen-2019-030538]
23. Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA). 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 [accessed 2019-09-18]
24. Thorndike AN, Riis J, Sonnenberg LM, Levy DE. Traffic-light labels and choice architecture: promoting healthy food choices. Am J Prev Med 2014;46(2):143-149 [FREE Full text] [doi: 10.1016/j.amepre.2013.10.002] [Medline: 24439347]
25. 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]
26. Dayan E, Bar-Hillel M. Nudge to nobesity II: menu positions influence food orders. Judgm Decis Mak 2011;6(4):333-342. [doi: 10.4324/9780203141939-11]
27. Glanz K, Bader MD, Iyer S. Retail grocery store marketing strategies and obesity: an integrative review. Am J Prev Med 2012;42(5):503-512. [doi: 10.1016/j.amepre.2012.01.013] [Medline: 22516491]
28. Schwartz M. The influence of a verbal prompt on school lunch fruit consumption: a pilot study. Int J Behav Nutr Phys Act 2007;4(1):6 [FREE Full text] [doi: 10.1186/1479-5868-4-6] [Medline: 17338812]
29. Just D, Price J. Default options, incentives and food choices: evidence from elementary-school children. Public Health Nutr 2013;16(12):2281-2288. [doi: 10.1017/s1368980013001468]

30. NSW Ministry of Health. The NSW Healthy School Canteen Strategy: Food and Drink Criteria. 2020. URL: https://education.nsw.gov.au/content/dam/main-education/en/home/student-wellbeing/whole-school-approach/healthy-school-canteens/202303_4th_edition_Food_Drink_Criteria_ACCESSIBLE.pdf [accessed 2019-09-18]
31. Xyris Software. Food Works. 2018. URL: https://xyris.com.au/whats-new-foodworks-9-professional/ [accessed 2021-10-30]
32. NSW Government. Healthy Food Finder. URL: https://www.foodfinder.health.nsw.gov.au [accessed 2021-10-30]
33. The George Institute for Global Health. FoodSwitch. URL: https://www.foodswitch.com.au/ [accessed 2021-10-30]
34. Australian Curriculum Assessment Reporting Authority. My School. 2019. URL: https://www.myschool.edu.au/ [accessed 2021-10-30]
35. Bucher T, Collins C, Rollo ME, McCaffrey TA, De Vlieger N, Van der Bend D, et al. Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. Br J Nutr 2016;115(12):2252-2263. [doi: 10.1017/s0007114516001653]
36. Skov LR, Lourenço S, Hansen GL, Mikkelsen BE, Schofield C. Choice architecture as a means to change eating behaviour in self-service settings: a systematic review. Obes Rev 2012;14(3):187-196. [doi: 10.1111/j.1467-789x.2012.01054.x]
37. Cochrane T, Davey R, de Castella FR. Estimates of the energy deficit required to reverse the trend in childhood obesity in Australian schoolchildren. Aust N Z J Public Health 2016;40(1):62-67 [FREE Full text] [doi: 10.1111/1753-6405.12474] [Medline: 26561382]
38. NSW Ministry of Health. NSW Healthy Eating and Active Living Stategy: Preventing Overweight and Obesity in New South Wales 2013-2018. North Sydney: NSW Ministry of Health; 2013.
39. Reilly KL, Reeves P, Deeming S, Yoong SL, Wolfenden L, Nathan N, et al. Economic analysis of three interventions of different intensity in improving school implementation of a government healthy canteen policy in Australia: costs, incremental and relative cost effectiveness. BMC Public Health 2018;18(1):378 [FREE Full text] [doi: 10.1186/s12889-018-5315-y] [Medline: 29558931]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ICC: intraclass correlation coefficient
NHMRC: National Health and Medical Research Council
NSW: New South Wales
OR: odds ratio
P&C: Parents and Citizens
P&F: Parents and Friends
RCT: randomized controlled trial
SEIFA: Socio-economic Indexes for Area

Edited by G Eysenbach; submitted 01.07.21; peer-reviewed by S Partridge, UK Chalwadi, M Salimi; comments to author 19.07.21; revised version received 08.09.21; accepted 26.09.21; published 29.11.21.

Please cite as:
Wyse R, Delaney T, Stacey F, Lecathelinais C, Ball K, Zoetemeyer R, Lamont H, Sutherland R, Nathan N, Wiggers JH, Wolfenden L Long-term Effectiveness of a Multistrategy Behavioral Intervention to Increase the Nutritional Quality of Primary School Students' Online Lunch Orders: 18-Month Follow-up of the Click & Crunch Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(11):e31734
URL: https://www.jmir.org/2021/11/e31734
doi:10.2196/31734
PMID:34847063

Original Paper

Views on Using Social Robots in Professional Caregiving: Content Analysis of a Scenario Method Workshop

Theresa Sophie Busse[1], MA; Sven Kernebeck[1], MScPH; Larissa Nef[2], BA; Patrick Rebacz[3,4], Dr med; Ilona Kickbusch[2], Prof Dr; Jan Peter Ehlers[1], Prof Dr med vet

[1]Department of Didactics and Educational Research in Healthcare, Faculty of Health, Witten/Herdecke University, Witten, Germany

[2]Careum Foundation, Zurich, Switzerland

[3]Visionom, Witten, Germany

[4]Department and Institute for Anatomy and Clinical Morphology, Faculty of Health, Witten/Herdecke University, Witten, Germany

Corresponding Author:
Theresa Sophie Busse, MA
Department of Didactics and Educational Research in Healthcare
Faculty of Health
Witten/Herdecke University
Alfred-Herrhausen-Straße 50
Witten, 58448
Germany
Phone: 49 2302 926 786 20
Email: theresa.busse@uni-wh.de

Abstract

Background: Interest in digital technologies in the health care sector is growing and can be a way to reduce the burden on professional caregivers while helping people to become more independent. Social robots are regarded as a special form of technology that can be usefully applied in professional caregiving with the potential to focus on interpersonal contact. While implementation is progressing slowly, a debate on the concepts and applications of social robots in future care is necessary.

Objective: In addition to existing studies with a focus on societal attitudes toward social robots, there is a need to understand the views of professional caregivers and patients. This study used desired future scenarios to collate the perspectives of experts and analyze the significance for developing the place of social robots in care.

Methods: In February 2020, an expert workshop was held with 88 participants (health professionals and educators; [PhD] students of medicine, health care, professional care, and technology; patient advocates; software developers; government representatives; and research fellows) from Austria, Germany, and Switzerland. Using the scenario methodology, the possibilities of analog professional care (Analog Care), fully robotic professional care (Robotic Care), teams of robots and professional caregivers (Deep Care), and professional caregivers supported by robots (Smart Care) were discussed. The scenarios were used as a stimulus for the development of ideas about future professional caregiving. The discussion was evaluated using qualitative content analysis.

Results: The majority of the experts were in favor of care in which people are supported by technology (Deep Care) and developed similar scenarios with a focus on dignity-centeredness. The discussions then focused on the steps necessary for its implementation, highlighting a strong need for the development of eHealth competence in society, a change in the training of professional caregivers, and cross-sectoral concepts. The experts also saw user acceptance as crucial to the use of robotics. This involves the acceptance of both professional caregivers and care recipients.

Conclusions: The literature review and subsequent workshop revealed how decision-making about the value of social robots depends on personal characteristics related to experience and values. There is therefore a strong need to recognize individual perspectives of care before social robots become an integrated part of care in the future.

(J Med Internet Res 2021;23(11):e20046) doi:10.2196/20046

KEYWORDS

social robots; robotics; health care sector; health personnel; ethics; forecasting; trends; technology; digital transformation; professional caregiving

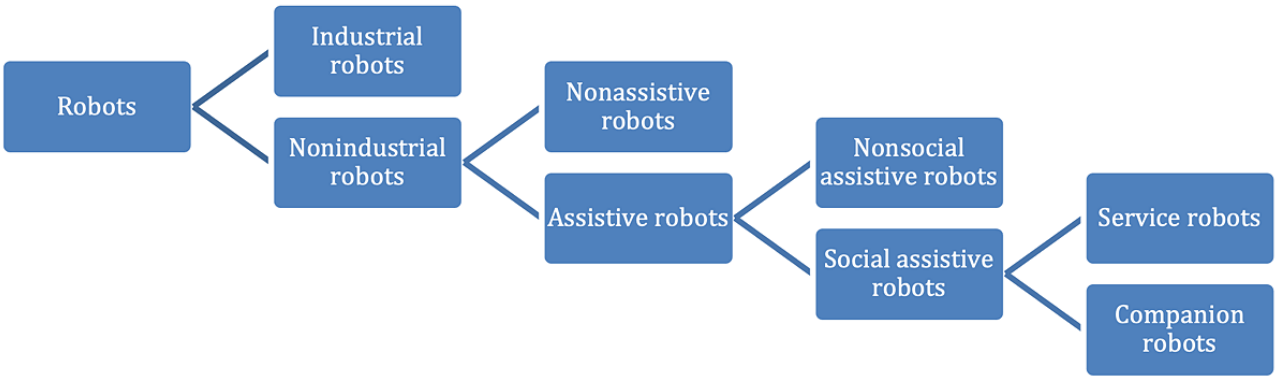
Introduction

Background

Social Robots in Care

Robotics represents a special case of technology use. Industrial robots are socially accepted to relieve humans from hard work [1]. We here focus on nonindustrial robots, which are relatively more complex than industrial robots and cover a broader range of applications being deployed in many different branches of commercial or private use [2] to assist people, including in health care. Nonindustrial robots can be divided into assistive and nonassistive robots. Assistive health robots can be used for surgery, therapy, or care [3], and can be subdivided into social or nonsocial assistive robots. Social assistive robots can be service robots (eg, a lifting aid) or companion robots (eg, animal-like entertainment robots) (see Figure 1) [4]. They interact with people or work closely with them. Social robots differ from pure service robots in that they emulate human behavior while providing services, thus establishing a form of interpersonal communication. For example, a socially acting lifting aid would not only reposition patients but also ask compassionately whether the lying position is comfortable. In this context, "interaction" does not necessarily have to take place via spoken language but can also take place exclusively via social and emotional cues [2]. Hereafter, the term "robot" is used to refer specifically to social robots.

Figure 1. Focused representation of the subdivision of the different robot types [4].



Robots are often researched in terms of their ability to express empathy and engage in interactive exchanges. Empathy can increase patient satisfaction and motivation for improvement as well as adherence to therapy programs in the patient-therapist interaction [5]. Robots simulate empathy mostly by facial and verbal expressions [6]. A small-scale study (N=36) found that people interacting with robots equipped with an "empathic module" communicated over a longer period of time and perceived the robots as more trustworthy, intelligent, and empathic than the control group who interacted with a robot that did not express empathy. In addition, the participants that interacted with the "empathetic" robot had a stronger sense of knowing the robot and perceived the interaction as comfortable [7]. However, another study showed that if there were incongruities between the affective state of the user and the emotional reaction of the robot, users rated the robot negatively [8]. A comparative study found that people were kinder to a robot that emulated empathy relative to the comparison group [4].

Growing Interest in the Use of Social Robots in Care

There are three core reasons for the growing interest in the use of robotics in care.

First, the demand for professional caregivers is likely to outstrip the supply [9]. Professional caregivers (ie, those working as trained specialists in the fields of health and nursing care or geriatric care in rehabilitation facilities, hospitals, and nursing homes, as well as in outpatient care in the home environment and related areas) are in demand in times of an aging population. In Switzerland, 367,000 people are being cared for by professional caregivers, which accounts for 28.9% of the population aged 80 or above; in 2018, 92,000 people, including 15.3% of the population older than 80 years, lived in a nursing home [10]. In the coming years, the availability of care is likely to decline as population growth slows and life expectancy increases. More people will need health care while fewer family members will be available to provide support. Due to the declining birth rate, future generations will consist of fewer people who could potentially take on the care of their parents or grandparents [11]. Other factors making it more difficult for family members to provide care include the rise in women's employment, with women traditionally taking on the caring role, and the growing mobility of future generations [11]. According to population forecasts, the percentage of people over 67 years of age will rise steeply [12]. The probability of neurodegenerative or chronic illness increases with age, which accordingly increases the care dependency of the population [13]. Assistive technologies may be one way to maintain care and ease this tense care situation. Robots may compensate for functional losses experienced by older people and promote everyday skills that help to maintain independence [2]. Robots can also enhance patient adherence to medications and exercise [14].

Second, robotics offers ways to increase the autonomy of service recipients. Elderly people are impaired in their autonomy if they are no longer able to perform certain activities independently, such as preparing food, housework, telephoning, and shopping.

For 16% of people living at home in Switzerland aged 65-79 years, one or more of these activities are either impossible or very difficult; 32% of people aged 80 years and over confirmed this statement, and minor difficulties were reported by 14% and 24% of the respondents aged 65-79 years and 80 years or older, respectively [15]. This shows that care often does not necessarily include physical care but that people also need support in areas that may be improved by technological assistance [3].

Third, robotics may take over the heavy work that puts a strain on professional caregivers. Three-quarters of professional caregivers state that heavy physical work is very common or frequent in their daily work [16]. Professional caregivers also report a high workload and associated stress. There is a mismatch between the workload and the time available [13]. Robots can relieve caregivers of the workload by supporting or even completely taking over heavy physical work tasks, thus saving time [3].

Acceptability of Social Robots

The development and potential of robots are still at an early stage. In 2018, 271,000 service robots were sold, representing a rise of 61% as compared with the market in 2017 [17]; however, implementation is progressing slowly in robotics as in other eHealth interventions [18]. The implementation raises many ethical questions such as data protection, responsibilities, or trust [19]. Many studies on barriers and facilitators have been published in recent years [20], indicating that concepts and applications should be widely debated by developers, experts, and users to ensure feasibility. Social acceptance is currently still low. For example, in the Eurobarometer survey, participants (N=26,751) generally had a positive attitude toward robots, but not in all areas: 27% stated that robotics should not be used in the health sector [21].

To maximize the benefits of using robots in health care, and especially in professional caregiving, it is necessary to learn about the attitudes and acceptance of these robots by society and health care professionals. A study has shown that people assess robotics as an opportunity. They expressed criticism of scenarios where robots performed care services for close relatives but liked the idea of robots undertaking dangerous tasks. The assessment of robots revealed only minor differences between different age groups, in which younger people were slightly less anxious about robots than the older or middle-aged groups [22]. Studies of perceived usefulness and perceived ease of use, which are critical to robot acceptance, are currently rare [23].

In speculating whether robots will have a role in providing care, we need to consider acceptance by both service recipients and professionals. This depends on the roles that robots undertake and the potential for expressing empathy [24,25]. The development of empathic robots is still at an early stage. To promote this development, we need to understand the factors that contribute to better care from the perspective of potential users of robotics. To determine whether robots are acceptable and how future caregiving can be designed, we organized a workshop with a series of focus group interviews among a diverse group of participants using future scenario planning.

Aim of the Study

This paper focuses on the development of a vision for the use of robotics in geriatric social and health care. It addresses the potential of social robotics to augment care for older people and supports the work of professional caregivers. This vision was developed during an expert workshop, including participants from Switzerland, Austria, and Germany, who have experience in the issues related to the care for older people. The workshop was based on four future scenarios, which served as inspiration for the participants' imagination regarding the possible changes in the field of professional caregiving through the use of social robots. The scenarios were developed in advance of the workshop using existing evidence from the literature. The workshop focused on potential and desirable future scenarios and on the steps required to prepare and implement these scenarios by 2025.

The aim of the workshop was to gather the views of health care professionals and educators, (PhD) students, patients, developers, scientists, and governments regarding the potential of robot use in future care and to stimulate debate and research on their use.

The research questions were: (1) Which (robotic) care scenario do the experts consider likely for 2025? (2) Which (robotic) care scenario do the experts consider desirable for 2025? (3) What can we do to make this scenario a reality?

Methods

Scenarios

An expert workshop (Careum Dialogue 2020) supported by the Careum Foundation was held in Zurich in February 2020. The workshop was designed according to the scenario method, which is suitable for statements about future goals with special consideration of the influencing factors and their effect on the goals. Consequences for future actions are to be derived from the scenario method. For this purpose, a best-case scenario and a worst-case scenario are formulated, which limit the range of conceivable developments and deviate from the long-term trend into positive or negative outcomes [26]. However, since the workshop participants had to discuss which scenarios represent the worst and best cases, two poles were formed: Analog Care (professional caregiving without any technical support) and Robotic Care (professional caregiving performed by robots without humans). In the sense of the scenario method, it is possible to define further characteristics between the poles of best-case and worst-case scenarios. However, the following criterion must be taken into account in the creation: the developments within the different scenarios must not cancel each other out. Furthermore, the scenarios must not be susceptible to collapse due to changes in minor factors. The extreme scenarios have the highest possible degree of severity and are thus as close as possible to the edges of the funnel of the possible future [26]. To support the experts in their discussion, two further scenarios were added: the scenario of professional caregivers supported by robotics (Deep Care) and the scenario in which both professional caregivers and robots work in teams and perform tasks independently (Smart Care). These scenarios are positioned between the two extreme

scenarios described above (Analog Care and Robotic Care). This was intended to improve understanding of the detailed different attitudes of the participants.

The scenarios used in this workshop were created with the help of literature collected by searching the PubMed and Google Scholar databases using the search terms “caregivers,” “care,” “robot,” “future,” “scenario,” “vision,” “utopia,” and “dystopia.” The development included all articles (N=28) dealing with the preferred care scenarios of the future, possibilities of robot use in care, ethical aspects of robot use in health care, changes in care through digitization, and the acceptance of robots (see Multimedia Appendix 1). Despite the wide range of scenarios, special attention was paid to developing the scenarios as “simple” scenarios. In the literature, “simple” scenarios are characterized in particular by few factors in their construction and description [27]. The reduced complexity in the description, even during the workshop, should allow a short time for the explanation of the scenario construct and create a larger scope for discussion. Thus, the scenario method enables the development of a spectrum of future visions. Participants were able to develop their own scenarios within this spectrum in the workshop, corresponding to their own ideas. The scenarios developed in advance served as support for creativity.

The scenarios were created according to the recommendations for scenario development by Fink and Siebe [27]. The addressees were experts from the health care sector in Switzerland, Germany, and Austria, who were to be supported in orienting the planning of future health care in politics, practice, and society. The workshop served as an orientation situation, which is distinct from a decision situation, and is not necessarily associated with a concrete decision between several alternatives for action but rather serves to orient and prepare for future decisions [27]. The scenarios were initially regionally limited to German-speaking countries, since the experts are employed in this field and regional particularities such as legal regulations or training guidelines are decisive when considering future developments in the health care system.

Data Collection

The data were collected by the experts using a flip chart exercise and detailed field notes of the group discussion.

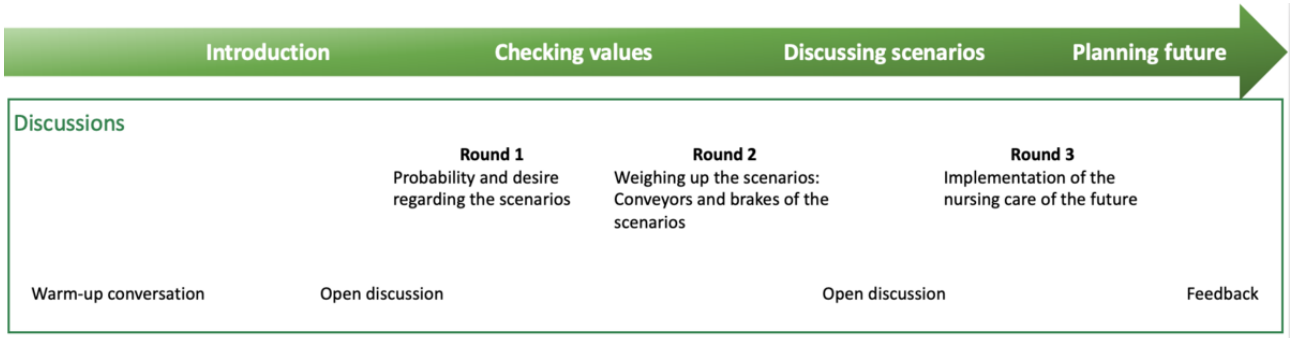
In the first round, one specific scenario was assigned to two tables each as the subject of debate. The advantages and disadvantages of Robotic Care, Analog Care, Smart Care, and Deep Care were addressed. The guiding questions for the first round of the group discussion were: (1) Has the scenario been described properly? (2) How likely is the scenario? (3) How desirable is the scenario?

In the second round, seating arrangements were changed according to plan so that experts from each scenario who sat at different tables beforehand now sat together. They briefly presented the discussion held at their table and began to talk about the likelihood and attractiveness of the different scenarios as well as the factors accelerating or limiting this health care future. Subsequently, the participants began to build their own health care scenarios such as a mix of the options given or a completely new alternative. The guiding questions for the second round of the group discussion were: (1) Which scenario is the most likely/desirable scenario? (2) Which aspects can accelerate or slow down the scenario? (3) Is there any other scenario you can imagine?

The groups at the eight tables reassembled again after this round. The last round focused on possible ways to compile and implement robotics in future professional caregiving. The question for the last group discussion was: Which steps are needed for implementation?

The procedure of the workshop is shown in Figure 2.

Figure 2. Workshop design.



To stimulate discussion, keynote speeches were given by experts between the various rounds. These dealt with the possible areas of application of robots, artificial intelligence, ethical requirements, the collaboration of robots with health care professionals, and the social interaction between robots and humans. Transcripts of the keynote speeches can be found in Multimedia Appendix 2.

Data Analysis

For evaluation, the research team’s records of the plenary discussions, the presentation of the respective discussions, as well as the sequential records of the discussions at the tables were analyzed using qualitative content analysis. The notes of the experts during the discussion at the tables were also included in the content analysis. The content analysis was carried out using MAXQDA line by line, sentence by sentence, or section by section according to the strategy of the inductive category formation described by Mayring [28]. With this approach, categories were formulated directly in the material from text passages that were evaluated according to the category definition. After a first step of category formation, some categories could be generalized step by step into main categories

by means of the summary. A total of 176 codes were identified (85 in the authors' notes and 91 in the participants' notes). These codes were organized into a common category system with five superordinate categories and a total of 25 subcategories and 21 subsubcategories. The following superordinate categories were formed: attitude toward robotics (25 codes), requirements for the care of the future (22 codes), discussion of scenarios (33 codes), implementation of professional caregiving in the future (20 codes), and development of professional caregiving in the future (76 codes).

The superordinate categories were based on the main components of the workshop: the recording of the general attitudes of the participants toward robotics, illumination of the needs for future care, discussion of the scenarios created in advance, as well as the development and implementation of the care in the future. The subcategories were also derived from the analyzed data according to the main focus of the workshop. For example, one subcategory contained all statements on the Robotic Care scenario. In turn, the subsubcategory dealt with different versions of the statements on the Robotic Care scenario, such as concern about the danger of two-tiered medicine. All statements made by participants that dealt with a concern in this respect were summarized in this subsubcategory.

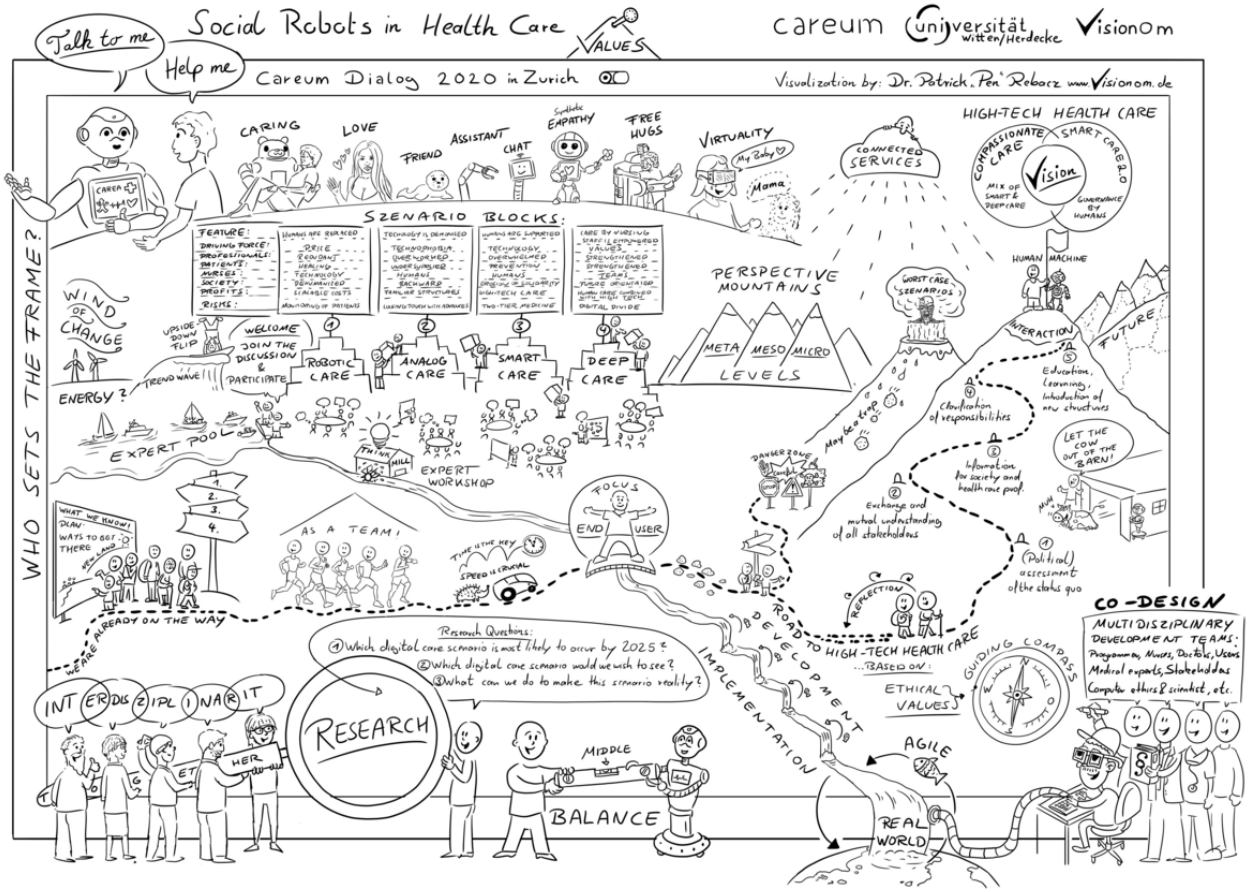
Ethical Aspects

Ethical principles of research in accordance with the Helsinki Declaration were observed [29]. The participants did not suffer any damage or impairments. The anonymity of the data was guaranteed by the form of data collection. Discussions were documented anonymously. Notes on flip charts were photographed and written down. The lecturers mentioned by name were asked for their consent. The participants had already been informed in the invitation about the focus of the workshop and the data collection planned. Participation was subject to consent.

Results

Figure 3 shows the various discussion contents of Careum Dialogue 2020 as a graphic recording. This graphic was presented in the follow-up workshop by one of the authors, who also took part in the discussion, based on the qualitative content analysis and serves here as a short overview before explicitly answering the research questions and presenting the further results.

Figure 3. Graphic recording of the Careum Dialogue 2020 workshop.



From the experts' point of view, either the occurrence of the Smart Care scenario (teams of professional caregivers and robots) or a mixture of the Smart Care and Deep Care scenarios (professional caregivers supported by robotics) were the most likely (robotic) care scenarios for 2025. From their point of view, such a development would be close to current health care and would enable professional caregivers to support patients without depriving them of the human contacts that are crucial for their well-being. The two extreme scenarios were felt to be unlikely. The experts saw neither sufficient technical nor social

prerequisites for the Robotics Care scenario to occur. In their view, the current state of technical development does not allow for such a complete takeover of care by robots. The experts also stated that not only technical progress was necessary to enable changes in the future but also the progress of society. For example, it would be necessary to implement training in eHealth literacy for the acceptance and critical evaluation of technology in the health care sector. The experts saw the occurrence of the Analog Care scenario as unrealistic owing to the limited human resources. In addition, they believed that a complete departure from technology would mean a step backward for both professional caregivers and patients, which would not be supported by society.

As part of the discussion on the most desirable (robotic) care scenario for 2025, the experts initially developed two scenarios on their own in small groups (Compassionate Care and Deep Care 2.0). These scenarios were based on the predeveloped Deep Care and Smart Care scenarios and supported the discussion about the various aspects of a desirable future scenario. For the future, the experts wished for the support of professional caregivers by technology and robotics. The experts believed that a combination of technology and professional caregivers is unbeatable, as technology could act on the latest evidence to support evidence-based care that leads to better results. In the experts' view, professional caregivers could thus possibly provide higher-quality care according to "traditional" values: safety can be increased by robotic support; caregivers can spend more time with patients by using robots to relieve them of routine tasks; patients' autonomy can be improved and extended by robotic support in their home environment; and patients can put together their own individual range of care services, which can consist of robotic and personal support. It was stated that patients might like the choice between robotics or professional caregivers, depending on their situation and the daily condition. In the debate, the experts mentioned that patients may perceive personal assistance as a burden at times. As an example, it was expounded that on some days, personal interaction and corresponding necessary courtesy are not pleasant. In addition, the experts reasoned that while health professionals may treat people differently depending on personal sympathies or other factors, robots might treat all patients equally.

The experts saw health care professionals as users whose acceptance is critical to the project. In addition to personal attitudes toward technology, the experts also considered education and the development of skills to be relevant for a high level of acceptance. Therefore, experts also advocated improving education and training to better prepare professional caregivers for this work. The long time needed for education and training while technology is changing so rapidly could be a problem in realizing this scenario. The experts saw the opportunity to enhance the value and reputation of health care through technology as the greatest benefit of this scenario. Professional caregivers could benefit from a redistribution of power as a result of the increasing use of digital tools. They could also benefit from a new professional profile that would attract people who are both highly socially oriented and technically interested. By uniting humans and technology, a two-tiered health care system could be avoided, which in the experts' view could develop through an increase in technical support. However, the experts were not sure whether the first-class health care would be robotic or personal health care.

The Robotic Care scenario was considered problematic because of the danger of two-tiered medicine and uncertainty about who would have the power to make decisions in this environment. The experts described care as highly individual and sensitive work, and saw the danger that standardized decisions could not reflect this complexity in all areas due to the exclusive use of robots. Another counterargument was the description of self-determination as the highest value. People should be able to refuse health care at any time. The experts expressed the fear that a robot could restrict self-determination if it offered standardized care that could not be controlled by the patients.

Further, it was discussed that patient acceptance is likely to have a large role in the adoption of robotics in health care. Similar to the case for health care professionals, the experts also saw the care recipients as users of robotics. The experts assumed that their acceptance depends not only on the implementation of robotics (eg, robotics enabling individualized care) but also on the area of application of the robotics and the scope of decision-making authority.

The Analog Care scenario was not considered desirable owing to its backward-looking nature and the resulting heavy burden on professional caregivers.

In the discussion on how to achieve the scenario they perceived as desirable, the experts saw a need for change in politics, institutions, health care professionals, science, health insurance funds, and society. The theory-practice-transfer route was urgently requested by the experts. In the experts' view, the hard road to high-tech health care based on ethical values therefore requires the following steps: (1) (political) assessment of the status quo; (2) exchange and mutual understanding of all stakeholders; (3) information for society and health care professionals; (4) clarification of responsibilities; and (5) education, learning, and introduction of new structures.

Within the framework of the analysis, the results were structured using the frequently used social science classification into three levels of analysis (macro, meso, micro) [30]. At the macro level, political systems or society as a whole are examined, whereas at the meso level, the focus is on parts of these systems such as institutions. At the micro level, individuals are analyzed along with their actions, decisions, and relationships.

At the macro level, the experts called for laws, norms, and regulations to be adapted to new needs. The financing of technical solutions and the setting of financial incentives for the further development of technical solutions must be made possible by political decision-makers. Education and training must also be comprehensively filled with new content at this level. For example, robotics and eHealth should be part of the curriculum. Such new guidelines could be developed by national groups with an international focus. The experts also saw a need for social change. A broad discussion on the desired care of the

future was needed with the aim of defining values that are important to society.

With regard to the meso level, institutions will have to introduce several changes for a positive health care scenario to succeed. The experts recommended bringing together information from Switzerland, Germany, and Austria to gain an overview. An exchange of information would create a network that could help to deal with the complexity of the sectors and the requirements of interoperability. Health care professionals should use this network as well as scientists. The experts also stated that institutions should set up experimental wards where health care professionals and developers can explore and test specific applications. This coordinated development with innovation labs and simulation centers could generate positive examples for practice and training. Moreover, cocreation was seen as the best way to develop new solutions for the future. Feedback between science and practical experience was seen as helpful to fine-tune possible solutions. In exploring the potential uses of robotics, experts considered the scientific study of the impact of robots on patient well-being and health as essential.

At the micro level, the experts saw a great need for empowerment of professional caregivers and patients. Professional caregivers would need to understand why changes in health care can enhance the value of their jobs. This can be achieved through sensitization and qualification, combined with participation and transparency. At several points in the discussion, the question arose as to which tasks could be performed by robots. In the experts' view, this indicates the need for a clearer definition of the professional caregivers' profession as distinguished from other professions.

Challenges at all levels were seen in users' unrealistic expectations, the possibility of collective standardization, and negative attitudes toward technology in the process of development. Communication might be another challenge. People of different professional groups need to find a mutual language to work together. The awareness of this fact is important for a successful work process.

Discussion

Principal Insights

During the workshop, the acceptance of robots, changes in the organization of work resulting from the use of robots, and new ethical and legal requirements were the main topics of discussion raised. These topics were also identified in a previous study as central challenges in the introduction of robotics [31]. The discussions were stimulated by the possible visions of the future. Ultimately, the experts spoke in favor of a future in which professional caregivers are supported by robots. The experts considered this scenario to be both the most desirable and the most probable.

The experts' opinions regarding the most probable and desirable future scenario are in accordance with the results of other international studies. First and foremost, a similar sentiment was expressed in a workshop with 25 Australian research or health care experts regarding future prospects on the subject of digital health. The scenarios developed by the participants, which represented their wishes and ideas for future care, included the use of robots as a support for older people to maintain independence and health (interaction, housework, transmission of health data) as well as the use of robots in the field of public health care and diagnostics [32]. In addition, it is important to consider that the visions of the future described by the experts in the workshop coincide with the desires of health care professionals and trainees that have been collected in studies. For example, in one study, professional caregivers most frequently named the desired use of robots in the areas of play, occupation, and activity; support for functional mobility; and in the supply/disposal of materials [33]. In another study, medical and nursing students (N=178) described a desire to use robots to remind the elderly to take medication, monitor their health, and promote physical and mental exercise [34].

The topic of acceptance was a particularly prominent focus of the discussion during the workshop. The experts believed that user acceptance is decisive for further implementation. In this case, the users are the professional caregivers but also the care recipients. From the experts' perspective, both groups should receive the necessary education and information to be able to assess and understand robotics and thus increase acceptance. Acceptance is also named in the literature as the decisive factor for the use of robotics [35] and is therefore discussed more intensively below.

The acceptance of robots must be considered separately for different user groups. First, we consider whether and how robots are accepted by the older population. A review of the use of robots in therapy and care found that robots were perceived positively by the older population [36]. They felt safer because the robots were able to detect emergency events. In addition, the potential of robots to improve social skills of the users or to alleviate loneliness and isolation was evaluated positively by older people. However, this positive acceptance of robotics is also accompanied by various fears: people named the fear of losing human contact, being deceived by robots with regard to their abilities, and the fear of infantilization by using robots as toys [36].

Second, the acceptance of users working together with robots is considered. For instance, in a review on the social acceptance of robots in different professional fields, 336 articles were extracted from four databases [37]. After consideration of the exclusion criteria, 39 articles remained to be included in the review. In general, the review for the health and social services sector indicated a much more positive basic attitude toward robots than in other sectors. However, the studies also found that people who had no experience with robots more often had a negative attitude toward robots [37].

Third, acceptance must also be considered in relation to the respective field of activity and the degree of decision-making power. The concerns of experts regarding automated decisions of robots that lead to care that is not value-congruent or not desired by patients could be similarly considered. Poulsen and Burmeister [38] tested a new framework for care robots: the robot should be able to provide value-sensitive, individually adapted patient care. The study investigated the willingness of the end users to trust the decisions of the artificial intelligence.

They described different scenarios of care for elderly people with robotic support, which were linked to personal values (autonomy, respect, dignity, privacy, independence, social connectedness) of the person. In the first phase of the study, the interviewed experts (N=4), including a registered nurse, a robotics academic, a computer ethicist, and a computer scientist, were able to see how the robot behaved differently in the same situation depending on which of the values was the most important for the person receiving care. The same scenario was shown several times, but with the supervised person giving a different rating of the personal values. For example, a person who was particularly concerned with autonomy would ask for help at a later time than a person who had rated this value as less important. Three of the four people subsequently stated that they considered the robot's care to be value-sensitive and that this would enable them to receive good and individual care. In the second phase of the study, subjects (N=102) were shown two of the scenarios from phase 1 in a slightly adapted form. After each value had been changed, a questionnaire was administered to assess whether the participants would accept the decisions of the robots and use them in their daily lives. In addition, it was asked how trustworthy the users rated the robot outside the scenarios. The participants rated the robot as trustworthy for the scenarios in the questionnaires (50/102 questionnaires were completely filled out). However, in considering the supply chain beyond the given scenarios, the majority of respondents (66%) were not prepared to trust the robot if they were not clear about how decisions are made. In addition, 82% of the respondents indicated that they would like to have the ability to change the way the robot makes decisions [38]. This shows that the willingness to trust a robot is strongly dependent on how comprehensible and controllable its decisions are.

In summary, from these studies, people generally have a positive attitude toward robots and want to interact with them. A review on the acceptance of robots supports this statement [39]. Nevertheless, the field of health care should be evaluated with high sensitivity and there is a need for further research to fully understand crucial factors for the acceptance of robots.

Conclusion

Health care that is characterized by the combination of robotics and humans has great potential to support the independence of care recipients, improve health outcomes, and relieve the burden for caregivers. There are already some approaches to support professional caregivers with technology. The results of this workshop show that technology-supported care is the care of the future favored by experts. To determine the exact characteristics of this type of care in the future, it is necessary to ascertain the wishes of society in the German-speaking countries in addition to the wishes of the experts. It is also necessary to adapt the legal regulations to create incentives for technical progress, legally define the necessary competencies for the professional caregivers, and implement them in education and training with the help of suitable teaching materials [40,41].

Future research should focus on what society, and in particular those in need of care, demand for the care of the future. However, the perspective of professional caregivers is also critical to development. One review pointed out that existing studies on health care workers' perceptions have mainly focused on the impact robots might have on their patients rather than on themselves [42]. Moreover, the development of assistive technologies within the framework of scientific projects should be carried out in multidisciplinary teams with the involvement of users in the sense of co-design [43-45]. The co-design process can help to increase the acceptance in society and thereby also of the users, and to develop the technology oriented to the needs of the users [46].

Thus, care recipients are enabled to use technical support in their everyday lives. In addition to the already existing research in the field of eHealth literacy and technology acceptance, it is crucial to develop approaches for the training and education of society adapted to the scenarios of the future and the respective settings.

Acknowledgments

This article is an output from a 2-day workshop on the ethically acceptable integration of robots in future professional caregiving settings with a focus on the challenges for stakeholders and end users that was held in February 2020. The workshop was supported by the Careum Foundation. The authors of this article attended the workshop. They agreed to develop the key statements and take the required further steps toward a peer-reviewed publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature on the basis of which the scenarios for the Careum Dialogue were developed.
[DOCX File , 20 KB - jmir_v23i11e20046_app1.docx]

Multimedia Appendix 2
Presentations held at Careum Dialogue 2020.
[PDF File (Adobe PDF File), 109812 KB - jmir_v23i11e20046_app2.pdf]

References

1. Jörg J. Künstliche Intelligenz und Robotermedizin. In: Digitalisierung in der Medizin. Wie Gesundheits-Apps, Telemedizin, künstliche Intelligenz und Robotik das Gesundheitswesen revolutionieren. Berlin, Heidelberg: Springer-Verlag; 2018:85-111.
2. Janowski K, Ritschel H, Lugrin B, André E. Sozial interagierende Roboter in der Pflege. In: Bendel O, editor. Pflegeroboter. Wiesbaden: Springer Gabler; Nov 22, 2018:63-87.
3. Bendel O. Roboter im Gesundheitsbereich. In: Bendel O, editor. Pflegeroboter. Wiesbaden: Springer Gabler; Nov 22, 2018:195-212.
4. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A. The influence of empathy in human–robot relations. Int J Hum-Comput Stud 2013 Mar;71(3):250-260 [FREE Full text] [doi: 10.1016/j.ijhcs.2012.09.005]
5. Tapus A, Tapus C, Matarić MJ. Hands-off therapist robot behavior adaptation to user personality for post-stroke rehabilitation therapy. 2007 Presented at: IEEE International Conference on Robotics and Automation; April 10-14, 2007; Rome, Italy URL: https://authors.library.caltech.edu/76725/1/04209308.pdf [doi: 10.1109/robot.2007.363544]
6. Paiva A, Leite I, Boukricha H, Wachsmuth I. Empathy in virtual agents and robots. ACM Trans Interact Intell Syst 2017 Oct 09;7(3):1-40. [doi: 10.1145/2912150]
7. Charrier L, Galdeano A, Cordier A, Lefort M. Empathy display influence on human-robot interactions: a pilot study. 2018 Presented at: International Conferenc on Intelligent Robots and Systems (IROS); October 1-5, 2019; Madrid, Spain URL: https://hal.archives-ouvertes.fr/hal-01887075
8. Cramer H, Goddijn J, Wielinga B, Evers V. Effects of (in)accurate empathy and situational valence on attitudes towards robots. : IEEE; 2010 Presented at: 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 2-5, 2010; Osaka, Japan URL: https://ieeexplore.ieee.org/document/5453224 [doi: 10.1109/hri.2010.5453224]
9. Schwinger A. Pflegepersonal heute und morgen. In: Jacobs K, Kuhlmey A, Greß S, Klauber J, Schwinger A, editors. Pflegereport 2019. Mehr Personal in der Langzeitpflege – aber woher?. Berlin, Heidelberg: Springer-Verlauf; 2020:3-21.
10. Gesundheit. Taschenstatistik 2019. Bundesamt für Statistik. Neuchâtel, Suisse; 2020 Feb 14. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit.assetdetail.11807264.html [accessed 2021-10-18]
11. Kehl C. Wege zu verantwortungsvoller Forschung und Entwicklung im Bereich der Pflegerobotik: Die ambivalente Rolle der Ethik. In: Bendel O, editor. Pflegeroboter. Wiesbaden: Springer Gabler; Nov 22, 2018:141-160.
12. Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung. Variante 1, Variante 2 und Variante 3 im Vergleich. Destatis. Statistisches Bundesamt. 2019 Jul 04. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/Tabellen/variante-1-2-3-altersgruppen.html [accessed 2021-10-18]
13. Wie haben sich Angebot und Inanspruchnahme der Gesundheitsversorgung verändert? Gesundheitsberichterstattung des Bundes. Gesundheit in Deutschland2015. Berlin: Robert Koch-Institut; 2015. URL: https://www.gbe-bund.de/pdf/gesber2015.pdf [accessed 2021-10-18]
14. Broadbent E, Garrett J, Jepsen N, Li Ogilvie V, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
15. Grad der Einschränkungen in den instrumentellen Alltagsaktivitäten (IADL). Schweizerische Gesundheitsbefragung 2017. Bundesamt für Statstik. 2018. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/grafiken.assetdetail.8066068.html [accessed 2021-10-18]
16. Schmucker R. Arbeitsbedingungen in Pflegeberufen. In: Jacobs K, Kuhlmey A, Greß S, Klauber J, Schwinger A, editors. Pflegereport 2019. Mehr Personal in der Langzeitpflege – aber woher?. Berlin, Heidelberg: Springer-Verlag; 2020:49-60.
17. Executive Summary World Robotics 2019. 14. International Federation of Robotics. 2019. URL: https://ifr.org/downloads/press2018/Executive_Summary_WR_Service_Robots_2019.pdf [accessed 2021-10-18]
18. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
19. Nwosu AC, Sturgeon B, McGlinchey T, Goodwin CD, Behera A, Mason S, et al. Robotic technology for palliative and supportive care: Strengths, weaknesses, opportunities and threats. Palliat Med 2019 Sep 28;33(8):1106-1113 [FREE Full text] [doi: 10.1177/0269216319857628] [Medline: 31250734]
20. Papadopoulos I, Koulouglioti C, Lazzarino R, Ali S. Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review. BMJ Open 2020 Jan 09;10(1):e033096 [FREE Full text] [doi: 10.1136/bmjopen-2019-033096] [Medline: 31924639]
21. Public Attitudes towards Robots. Special Eurobarometer 382.: European Commission; 2021 Sep. URL: https://www.ab.gov.tr/files/ardb/evt/Public_attitudes_toward_robots_2012.pdf [accessed 2021-10-18]
22. Backonja U, Hall AK, Painter I, Kneale L, Lazar A, Cakmak M, et al. Comfort and attitudes towards robots among young, middle-aged, and older adults: a cross-sectional study. J Nurs Scholarsh 2018 Nov 19;50(6):623-633 [FREE Full text] [doi: 10.1111/jnu.12430] [Medline: 30230692]
23. Parviainen J, Turja T, Aerschot L. Social robots and human touch in care: the perceived usefulness of robot assistance among healthcare professionals. In: Korn O, editor. Social robots: technological, societal and ethical aspects of human-robot interaction. Cham, Switzerland: Springer Nature; 2019:187-204.

24. Saadatzi MN, Logsdon MC, Abubakar S, Das S, Jankoski P, Mitchell H, et al. Acceptability of using a robotic nursing assistant in health care environments: experimental pilot study. J Med Internet Res 2020 Nov 12;22(11):e17509 [FREE Full text] [doi: 10.2196/17509] [Medline: 33180024]
25. Vänni KJ, Salin SE. Attitudes of professionals toward the need for assistive and social robots in the healthcare sector. In: Korn O, editor. Social robots: technological, societal and ethical aspects of human-robot interaction. Cham, Switzerland: Springer Nature; 2019:205-236.
26. Helmrich R, Zika G. Prognosen, projektionen und szenarien. In: Baur N, Blasius J, editors. Handbuch methoden der empirischen sozialforschung. Wiesbaden: Springer Fachmedien; 2019:231-246.
27. Fink A, Siebe A. Szenario entwicklung. In: Szenario management. Von strategischem vorausdenken zu zukunftsrobusten Entscheidungen. Frankfurt am Main: Campus Verlag GmbH; Oct 13, 2016:63-123.
28. Mayring P, Fenzl T. Qualitative inhaltsanalyse. In: Baur N, Blasius J, editors. Handbuch methoden der empirischen sozialforschung. Wiesbaden: Springer; Feb 15, 2019:633-648.
29. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
30. Schubert K, Klein M. Makro-/Meso-/Mikroebene. Das Politlexikon. Begriffe, Fakten, Zusammenhänge. Bonn: Bundeszentrale für politische Bildung; 2020. URL: https://www.bpb.de/nachschlagen/lexika/politiklexikon/296425/makro-meso-mikroebene [accessed 2021-11-04]
31. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
32. Lupton D. Digital health now and in the future: Findings from a participatory design stakeholder workshop. Digit Health 2017;3:2055207617740018 [FREE Full text] [doi: 10.1177/2055207617740018] [Medline: 29942616]
33. Gliesche P, Seibert K, Kowalski C, Domhoff D, Pfingsthorn M, Wolf-Ostermann K, et al. Robotic assistance in nursing care: survey on challenges and scenarios. Int J Biomed Biol Engineer 2020;14(9):257-262 [FREE Full text]
34. Łukasik S, Tobis S, Kropińska S, Suwalska A. Role of assistive robots in the care of older people: survey study among medical and nursing students. J Med Internet Res 2020 Aug 12;22(8):e18003 [FREE Full text] [doi: 10.2196/18003] [Medline: 32784187]
35. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res 2018 Sep 21;20(9):e264 [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]
36. Cifuentes CA, Pinto MJ, Céspedes N, Múnera M. Social robots in therapy and care. Curr Robot Rep 2020 Jun 29;1(3):59-74. [doi: 10.1007/s43154-020-00009-2]
37. Savela N, Turja T, Oksanen A. Social acceptance of robots in different occupational fields: a systematic literature review. Int J of Soc Robotics 2017 Dec 6;10(4):493-502. [doi: 10.1007/s12369-017-0452-5]
38. Poulsen A, Burmeister OK. Overcoming carer shortages with care robots: Dynamic value trade-offs in run-time. Australas J Inf Syst 2019 Feb 12;23:1-19. [doi: 10.3127/ajis.v23i0.1688]
39. Naneva S, Sarda Gou M, Webb TL, Prescott TJ. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. Int J of Soc Robotics 2020 Jun 03;12(6):1179-1201. [doi: 10.1007/s12369-020-00659-4]
40. McAllister M, Kellenbourn K, Wood D. The robots are here, but are nurse educators prepared? Collegian 2021 Apr;28(2):230-235. [doi: 10.1016/j.colegn.2020.07.005]
41. Pepito JA, Locsin R. Can nurses remain relevant in a technologically advanced future? Int J Nurs Sci 2019 Jan 10;6(1):106-110 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.013] [Medline: 31406875]
42. Papadopoulos I, Koulouglioti C, Ali S. Views of nurses and other health and social care workers on the use of assistive humanoid and animal-like robots in health and social care: a scoping review. Contemp Nurse 2018;54(4-5):425-442. [doi: 10.1080/10376178.2018.1519374] [Medline: 30200824]
43. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
44. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
45. Van Velsen L, Wentzel J, Van Gemert-Pijnen E. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
46. Tuisku O, Pekkarinen S, Hennala L, Melkas H. "Robots do not replace a nurse with a beating heart". Inf Technol People 2019 Feb 04;32(1):47-67. [doi: 10.1108/itp-06-2018-0277]

Edited by R Kukafka; submitted 10.05.20; peer-reviewed by K Davies, M Dohan; comments to author 25.05.20; revised version received 21.06.20; accepted 23.09.21; published 10.11.21.

Please cite as:
Busse TS, Kernebeck S, Nef L, Rebacz P, Kickbusch I, Ehlers JP
Views on Using Social Robots in Professional Caregiving: Content Analysis of a Scenario Method Workshop
J Med Internet Res 2021;23(11):e20046
URL: https://www.jmir.org/2021/11/e20046
doi:10.2196/20046
PMID:34757318

Review

Blockchain Integration With Digital Technology and the Future of Health Care Ecosystems: Systematic Review

Hanaa Fatoum[1], MBBS; Sam Hanna[2], PhD, MBA; John D Halamka[3], MD, MS; Douglas C Sicker[4], PhD; Peter Spangenberg[5], MD; Shahrukh K Hashmi[6], MD, MPH

[1]College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

[2]School of Professional & Extended Studies, American University, Washington, DC, United States

[3]Mayo Clinic Platform, Mayo Clinic, Rochester, MN, United States

[4]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

[5]King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

[6]Division of Hematology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Hanaa Fatoum, MBBS
College of Medicine
Alfaisal University
Takhassusi
Riyadh, 11533
Saudi Arabia
Phone: 966 000000
Email: Hanaa.Fatoum@outlook.com

Abstract

Background: In the era of big data, artificial intelligence (AI), and the Internet of Things (IoT), digital data have become essential for our everyday functioning and in health care services. The sensitive nature of health care data presents several crucial issues such as privacy, security, interoperability, and reliability that must be addressed in any health care data management system. However, most of the current health care systems are still facing major obstacles and are lacking in some of these areas. This is where decentralized, secure, and scalable databases, most notably blockchains, play critical roles in addressing these requirements without compromising security, thereby attracting considerable interest within the health care community. A blockchain can be maintained and widely distributed using a large network of nodes, mostly computers, each of which stores a full replica of the data. A blockchain protocol is a set of predefined rules or procedures that govern how the nodes interact with the network, view, verify, and add data to the ledger.

Objective: In this article, we aim to explore blockchain technology, its framework, current applications, and integration with other innovations, as well as opportunities in diverse areas of health care and clinical research, in addition to clarifying its future impact on the health care ecosystem. We also elucidate 2 case studies to instantiate the potential role of blockchains in health care.

Methods: To identify related existing work, terms based on Medical Subject Headings were used. We included studies focusing mainly on health care and clinical research and developed a functional framework for implementation and testing with data. The literature sources for this systematic review were PubMed, Medline, and the Cochrane library, in addition to a preliminary search of IEEE Xplore.

Results: The included studies demonstrated multiple framework designs and various implementations in health care including chronic disease diagnosis, management, monitoring, and evaluation. We found that blockchains exhibit many promising applications in clinical trial management such as smart-contract application, participant-controlled data access, trustless protocols, and data validity. Electronic health records (EHRs), patient-centered interoperability, remote patient monitoring, and clinical trial data management were found to be major areas for blockchain usage, which can become a key catalyst for health care innovations.

Conclusions: The potential benefits of blockchains are limitless; however, concrete data on long-term clinical outcomes based on blockchains powered and supplemented by AI and IoT are yet to be obtained. Nonetheless, implementing blockchains as a novel way to integrate EHRs nationwide and manage common clinical problems in an algorithmic fashion has the potential for improving patient outcomes, health care experiences, as well as the overall health and well-being of individuals.

(J Med Internet Res 2021;23(11):e19846) doi:10.2196/19846

KEYWORDS

blockchain, Internet of Things; digital; artificial intelligence; machine learning; eHealth; ledger; distributed ledger technology

Introduction

The blockchain concept was first described 3 decades ago and was meant to be used as a digital timestamp for documents to prevent tampering, functioning somewhat like a notary. However, it did not develop significantly and went largely unnoticed until the global financial crisis. In 2008, the blockchain revolution began when Nakamoto pioneered and crystallized it by releasing his whitepaper [1] followed by his cryptocurrency called Bitcoin, offered as an open-access protocol to the public.

Blockchain is considered one of today's important ground-breaking technologies. The question is what makes blockchain so unique and useful.

In simple words, it provides digital trust and transparency, something that has not only been seriously lacking in the digital world but is also posing major challenges in an era of increased dependence on electronic data, along with viewing the shift toward digitization and substituting other traditional methods of data storage as a glorified goal. A caveat that needs to be considered is that many digital health start-ups have failed, given their inability to convince the investors and users or because of using older technologies that become outdated by the time a completely digital health system is established.

The security of digital data, and the fact that it can be manipulated, tampered with, and purposefully hidden to suit the parties of interest, can be quite an alarming thought. Security concerns have led to resistance toward the use of electronic cloud-based data. However, using blockchain technology can potentially provide a breakthrough.

We will evaluate cases involving 2 patients in typical yet different clinical scenarios and then analyze the role of blockchains in the management of these patients.

Chenoa is an 8-year-old girl from Cheyenne, Wyoming, who was recently diagnosed with acute lymphoblastic leukemia. Her parents were farmers and there was no large tertiary care center in town that could provide treatment for life-threatening cancers with aggressive chemotherapy. They traveled to Colorado to obtain specialized care as recommended by their local hematologist. She received all her initial treatments in Denver, followed by a stem cell transplant from her elder brother as the donor after receiving massive radiation and chemotherapy for the transplant. Moreover, 4 months posttransplant, there was no evidence of leukemia or rejection (graft-versus-host disease), and they return to Cheyenne to celebrate their cancer conquest with the rest of the Cherokee tribe. A month after returning, Chenoa develops high-grade fever and is diagnosed with an extremely low white blood cell count along with a relapse of her leukemia at the local hospital. Immediate transfer to the transplant center in Colorado is recommended.

Khaled is a 58-year-old retired banker and an ex-smoker who lives in Dearborn, Michigan, with hypertension, coronary heart disease, and chronic kidney disease for which he undergoes hemodialysis thrice a week and is on the renal transplant list for a transplant. He is divorced and single; however, his daughter, who lives in Cleveland (Ohio), frequently visits him. He was recently diagnosed with heart failure; given the precarious health system and lack of caretakers in Dearborn, he is temporarily planning to move to his daughter's house.

The above 2 examples characterize the real-world scenarios within the United States, a developed country; despite incurring some of the highest health care costs, the United States has a disjointed health care system as far as digital health is concerned. Given the impediments and complexities in the US health care system, errors and wastage within the health ecosystem can directly affect patient outcomes like those of Chenoa and Khaled.

Chenoa's transfer was delayed given that she became increasingly unstable, as her infection led to septic shock. Though intensive care unit (ICU) management was optimum, and she was receiving multiple antibiotics and vasopressors, hopes for her stabilization and transfer to Denver were diminishing. A couple of days prior to hospital admission, Chenoa was taken to another local hospital because of fever and flu-like symptoms, where she was given an outpatient prescription of antihistamines. Though her transplant team had instructed the parents to telephone them for any issues, this was a very minor issue, and not entirely unexpectedly, the local physician in the community did not realize the depth of the immunosuppressive state that Chenoa exhibited. A week after highly aggressive treatment in the ICU, Chenoa died of multiple organ failure due to sepsis.

Khaled's story had a different twist; his daughter made a hasty decision to take him to Ohio, but he deteriorated quickly. She had to take him directly to a large tertiary care center in Cleveland where the emergency room doctors evaluated and triaged him appropriately and were relentlessly trying to obtain medical records from Detroit where Khaled's doctors were located. He went into cardiac arrest twice within a few hours of arrival and cardiopulmonary resuscitation was stopped after a prolonged effort following discussions with his daughter, as it seemed futile.

Fortunately, most of the patients in the United States do not exemplify the above cases; however, similar issues routinely occur given the lack of trustworthy and secure digital infrastructure for health care. Many questions arise after the death of Chenoa and Khaled, the most essential of which is whether there was a medical error on the side of the individuals or the health care ecosystem, which led to their fatal demise (eg, were these deaths preventable?).

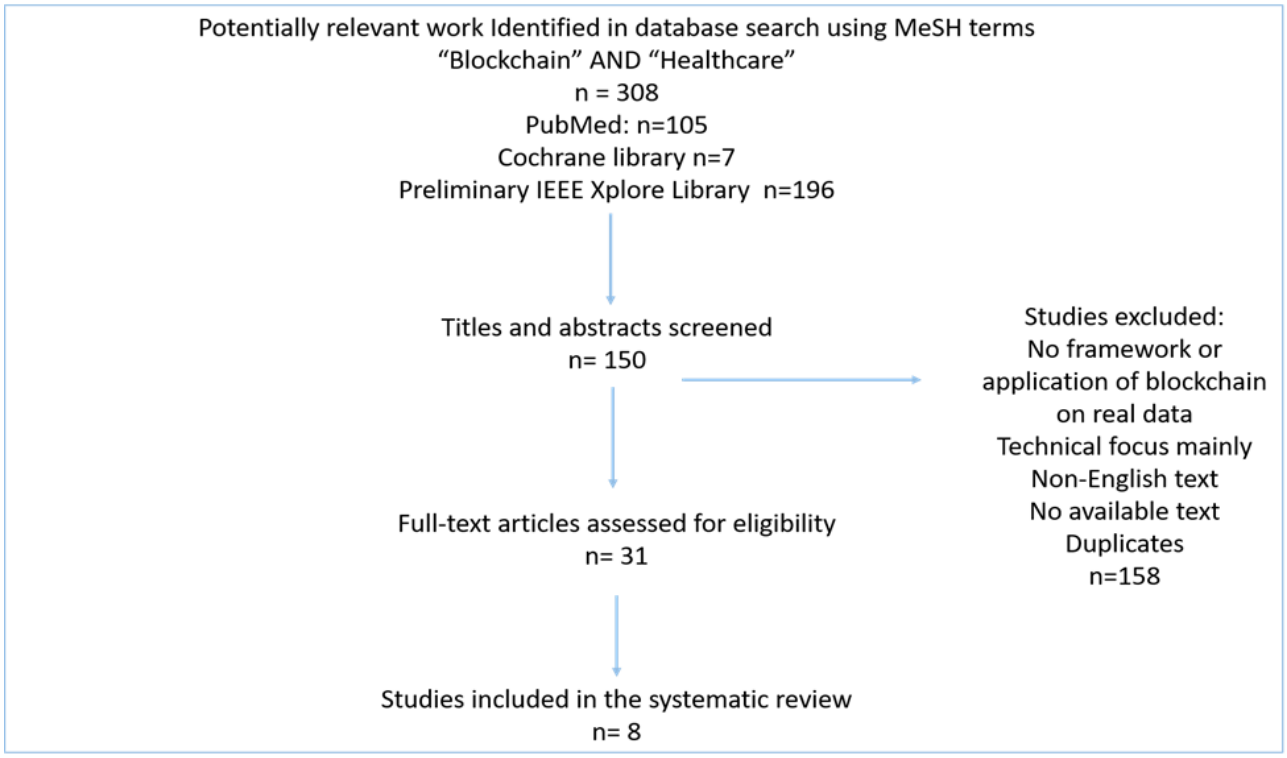
We provide a systematic review of literature on the novel health care ecosystem focusing on blockchains and then consider the cases of Chenoa and Khaled based on the current data.

Methods

Literature Search

The authors searched through the literature using the terms "blockchain" and "healthcare" based on Medical Subject Headings to identify related existing work. The inclusion criteria were focused on studies related to health care and clinical research; a working framework was proposed as well as implementation and testing with data. Exclusion criteria are presented in Figure 1.

Figure 1. Systematic review outcome. MeSH: Medical Subject Headings, IEEE: Institute of Electrical and Electronics Engineers.



The literature sources for this systematic review were PubMed, Medline, and the Cochrane library in addition to a preliminary search of IEEE Xplore, which had works with a technical focus rather than a health care outcome focus.

Reviewers created a data extraction sheet and identified the required data, and 31 full-text studies were critically evaluated; among these, 9 studies were finally selected, and they are summarized in Table 1.

First, we provide an overview of the essential elements of blockchains for readers and then explain the specific methodology in detail.

Overview of Blockchain Technology

The 3 pillars of blockchain technology that make it a revolutionary technology are decentralization, transparency, and immutability. A block is a virtual data storage unit that holds records, transactions, or other means of data. Each block is chronologically linked to the previous blocks using cryptography; once a block is created, it is permanently stored on the blockchain and cannot be modified or removed, thus becoming immutable. If one wishes to update a network node, make another transaction, or add new data, a new block must be added to the chain; all the previous blocks will still be unchanged and visible. Another aspect of a blockchain is that it is a decentralized database using the distributed ledger technology, which simultaneously stores a full replica of the data on multiple nodes, unlike most other data management technologies where data storage is centralized, meaning that it is stored at a single location, mostly on a single server or mainframe computer [2].

The immutability of a blockchain is attributed to different factors; each block has an autogenerated header, which has a unique identification and a timestamp. The block also has a hash key, which is the header of the previous block. Therefore, all blocks are interlocked within their respective blockchains.

Once a verified permissioned node (node C) wants to add new data or make a transaction, a request is sent to all associated blockchain nodes to verify that the content of node C's blockchain matches the content of all the other nodes (nodes A, B, etc). In addition to ensuring a complete match of all the block headers, a unique signature header will be generated for the new block. The new block is added to the blockchain only after the approval of the node.

Blockchains may be permissionless (public), permissioned (private), or sometimes hybrid, as described below. Public

blockchains allow any user to join, view, and add data to the ledger, which offers maximal transparency. However, private blockchains allow only those with access to interact with the network and are usually controlled and maintained by a single organization offering superior privacy and scalability compared to public blockchains. Finally, in consortium-based blockchains, which are hybrids of the other 2 types, the configuration for viewing and writing access is determined by a group of organizations or entities.

Based on this general overview of blockchain technology, we present the salient features of blockchains, which include cybersecurity, applications, and various domains associated with them.

Results

The results of this systematic review alluded to a wide variety of applications for blockchains and various technical differences in the methods used to establish a blockchain [3-10]; a summary is provided in Table 1.

All the studies pertained to electronic health records (EHRs), clinical trials [6-8], or device integration using Internet of Things (IoT) [3,5,9,10].

Researchers at the Chinese Institute of Physical Science have proposed a model called Med-PPPHIS, which combines permissionless and permissioned blockchains, aiming at a closed-loop method for chronic disease management; the authors used Med-PPPHIS for national physique monitoring and scientific exercise guidance using various self-invented IoT medical devices such as health parameter assessment tools, athletic and functional performance assessment devices, wearable heart rate monitors, and intelligent fitness equipment. The blockchain was tested by 25 virtual machine simulations over 500 nodes, with results revealing superior security, higher data transmission rates, and low latency [3]. Another method was used by Hylock and Zeng et al, in which they used a proof-of-concept tool to extensively test all the 16 configurations of their proposed framework in a variety of scenarios, and they demonstrated results similar to the above study [4].

The blockchain was integrated with IoT devices for evaluating and monitoring essential tremor disease; herein, patients were able to use their smartphones to report their location and activity, self-evaluate their disease activity, and log aggravating and relieving factors, in addition to the data from their smartwatches and multiple air systems, providing a holistic view of their disease status. The authors concluded that blockchains resulted in increased efficiency, scalability, decreased cost, and flexibility in data access management [5]. Other studies have employed IoT devices and blockchains for patient monitoring [9,10]. They have created systems that can analyze and manage medical sensor data as well as send alerts based on patient-customized threshold values and abnormal patterns using smart contracts while simultaneously integrating the data into the patients' EHRs.

The last 3 studies [6-8] focused more on the application of blockchains in biomedical research and clinical trials providing proof-of-concept frameworks featuring customized smart contracts, which allowed for more control over data access depending on researcher privilege levels and patient-controlled authorization. Researchers have also incorporated additional security measures such as biometric verification for physician access. Additionally, blockchains have been used as a solution for continuous trial monitoring by clinics, financial sponsors, and participants.

Blockchains are instrumental in many clinical and research domains. Therefore, some governments are exploring the option of having a nationwide blockchain for the EHRs of all their citizens. Estonia is the first country in the world with a digital health care ecosystem for EHRs based on blockchain technology and has provided the world with a model for data integrity and efficiency. Some other countries are in the process of adopting blockchains at the macroscopic level to reduce health care waste, increase efficiency, and ultimately improve outcomes. However, besides the conventional risks of technological failure, scalability remains a challenge; for instance, conducting 800 transactions per second for hyperledgers using blockchains to store continuous telemetry data is not yet practical. Nonetheless, researchers hope that with the current advances in technology, specific solutions for storage optimization and redesigning of blockchain will be available soon [11]. One additional study [12] evaluated the performance of a blockchain-based online machine learning tool available on the internet called ExplorerChain, which uses 3 separate and different data sets (myocardial infarction, cancer biomarkers, and length of hospitalization). The study concluded that the performance of ExplorerChain was as good as the central server–based algorithm while providing the benefits of a distributed model. Nevertheless, the tradeoff with some of those benefits was the cost of efficiency. However, with the rise of supercomputers, the costs associated with running a blockchain are likely to decrease over time.

Let us return to the unfortunate cases of Khaled and Chenoa. Imagine that they lived in a digitalized nation (ie, a smart country as opposed to a smart city), where all the EHRs were on a private or a consortium-based blockchain.

Table 1. Summary of data extraction results.

Reference	Title	Platform or model	Implementation	Features	Method (blockchain interface and IoT[a] device)
Zhou et al [3]	Med-PPPHIS: blockchain-based personal healthcare information system for national physique monitoring and scientific exercise guiding	Med-PPPHIS and Med-DLattice blockchain	Chronic disease management: physique monitoring	Chronic disease management target; scientific and personalized exercise prescriptions (electronic prescriptions), providing users with safe, effective, and private scientific health guidance for the management of chronic diseases	Web portal and mostly self-developed medical IoT devices such as health sign monitoring equipment, heart rate monitor, and intelligent fitness equipment
Hylock and Zeng [4]	A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study	HealthChain	Patient-centered blockchain framework	Smart contracts, proxy re-encryption, revocable access, and patient-centered framework	Web portal
Zheng et al [5]	Accelerating health data sharing: a solution based on the Internet of Things and distributed ledger technologies	IOTA tangle	Remote diagnosis of essential tremor disease	App allowing users to report their location, activity, and tremor level; self-evaluation of the disease and other factors related to the disease, such as medication and alcohol consumption	App; wearable devices (Pebble smartwatch) and stationary air quality sensors
Zhuang et al [6]	Applying blockchain technology for health information exchange and persistent monitoring for clinical trials		Patient-reported outcomes for trials; EHR[b] sharing	Private blockchain, smart contracts, biometric verification of physicians, and patient-controlled data access	Web portal
Johnson et al [7]	Building a secure biomedical data sharing decentralized app (DApp): tutorial	Oasis Devnet/Ethereum	Biomedical research	Smart contracts, public code for app recreation, and geolocation sharing	iPhone (iOS) app
Maslove et al [8]	Using blockchain technology to manage clinical trials data: a proof-of-concept study	BlockTrial/ Ethereum protocol	Clinical trials	Smart contracts and mediated data access based on patient-granted permissions	Web app
Satamraju and Malarkodi [9]	Proof of concept of scalable integration of Internet of Things and blockchain in healthcare	DApp/Ethereum protocol	Remote patient monitoring	Smart contracts, off-chain storage, and Ethereum-based system users including patients, doctors, pharmacists, and insurance companies	App; pulse oximetry device, body temperature sensor, and room temperature sensor
Griggs et al [10]	Healthcare blockchain system using smart contracts for secure automated remote patient monitoring	DApp /Ethereum protocol	Automated patient monitoring	Smart contracts, custom threshold values for alerts, Oracle (master device to control smart contacts), and integration with EHRs	App; medical IoT devices
Kuo et al [11]	Expectation propagation logistic regression on permissioned blockchain (ExplorerChain): decentralized online healthcare/genomics predictive model learning	ExplorerChain	Blockchain combined with artificial intelligence for health care and genomics; predictive model training	iDASH private HIPAA[c]-compliant computing environment network; applied on 3 different data sets including myocardial infarction, cancer biomarkers, and length of hospitalization	Distributed servers using internet-based machine learning

[a]IoT: Internet of Things.

[b]EHR: electronic health record.

[c]HIPAA: The Health Insurance Portability and Accountability Act of 1996.

Discussion

Cybersecurity

EHRs play a vital role in providing smooth, safe, and efficient health care delivery. As many countries do not currently possess a unified health care system, data sharing and interoperability (in a secure manner) become essential for providing patients with the best care. However, this has been a major issue in current health care management systems. Data are one of the most valuable commodities; since the increased reliance on digital systems, cybercriminals have adjusted their methods owing to huge financial incentives, especially through selling the identifying data of people. Thus, medical records are currently worth more than social security numbers on the black market, as they include the date of birth, home addresses, contact data, health data, and other sensitive data. EHR data can be employed for various criminal activities like identity frauds, insurance frauds, phishing, and ransomware, leaving the patients compromised and susceptible to harm [13]. Cyberattacks on hospital systems have increased worldwide in the past decade; such attacks not only impede health care delivery and cause financial losses but also affect patients' trust in medical providers [13,14].

In 2018, a cyberattack on SingHealth (Singapore Health Services) compromised the records of 1.5 million patients [15]; it is considered one of the biggest data breaches in Singapore and worldwide. It has stirred considerable controversy about the security of patients' data and the reasons behind not implementing any significant measures for changing the EHR data systems. Compared to the conventional methods of EHR storage, a blockchain is a potentially secure and an immutable method for data storage and management owing to its decentralized nature. This means that the latest version of the chain is replicated, sent, and widely distributed in a huge network of nodes; there are no weak links for hackers to breach. Each block has a key of its own, in addition to having the hash key of the previous block. Once a transaction request has been made by a user, the blockchain protocol requests the network nodes to verify the validity of the entire blockchain content. This in turn means that unless the networks nodes verify that the current version of the chain is identical to theirs and approve the transaction, it cannot be added to the chain [16,17].

Furthermore, the transaction validation uses cryptominers, which are nodes that possess specialized hardware and software capable of solving energy-intensive cryptographic puzzles. It would be extremely difficult for someone to gather enough computing power to hack the blockchain database by altering the ledgers. The larger the blockchain, the more distributed it is, the more enormous is the computing power required for hacking, and the more secure it becomes [16,17].

Finally, digital signatures are employed to verify the identity of those who wish to access or add data to the blockchain. Additional features such as hardware security modules (HSMs) can be added as an additional layer to further enhance the protection of the patients' data. HSMs are specialized hardware devices that are used to guard highly sensitive data. An HSM acts as a trusted network node that performs several cryptographic processes such as key generation and management, as well as encryption and decryption of digital signatures. HSMs are usually placed in a secure physical location and cannot be accessed, thereby making them highly tamper-resistant systems [18].

Applications and Domains of Use in Health Care

Blockchain technology has thrived in many industries ranging from banking to supply chain management. It is predicted to have a major impact on the health care industry. According to the forecast report for 2018 to 2023 provided by Market Research Future, the global blockchain market is predicted to expand exponentially.

Fundamentally, from a patient's perspective, the potential role of a blockchain in developing a patient's personal health record could be significant. The patient-facing applications of this technology would benefit from one window and one operating system for the personal health records, out-of-pocket costs, covered versus uncovered services, clinical trial searches, consenting for clinical trials, and "omics" data interpretation.

The review focuses on 5 main areas in which most health care–related implementations fall under remote patient monitoring (including IoT devices), EHRs, patient-centered interoperability, clinical trial data management, and monetization, as shown in Table 2 and Figure 2. The potential application areas of blockchain technology in health care are depicted in Figure 3.

Table 2. Areas of blockchain implementation.

Application area	Blockchain technology features
Electronic health record management	• Patient-reported outcomes • Consent
Patient-centered interoperability	• DNR[a] orders • Instantaneous data access and interoperability
Remote patient monitoring	• Patient-mediated and controlled record access • IOT[b]-enabled monitoring of vital signs, glucose, and other parameters • Disease surveillance and outbreak management
Clinical trial data management	• Increased RCT[c] data transparency and quality • IOT-generated clinical research data • Smart contracts applying data specifications and incentives
Monetization	• Revenue cycle management • Clinical trial budgets

[a]DNR: do not resuscitate.

[b]IoT: Internet of Things.

[c]RCT: randomized control trial.

Figure 2. Illustration of blockchain usage in health care.

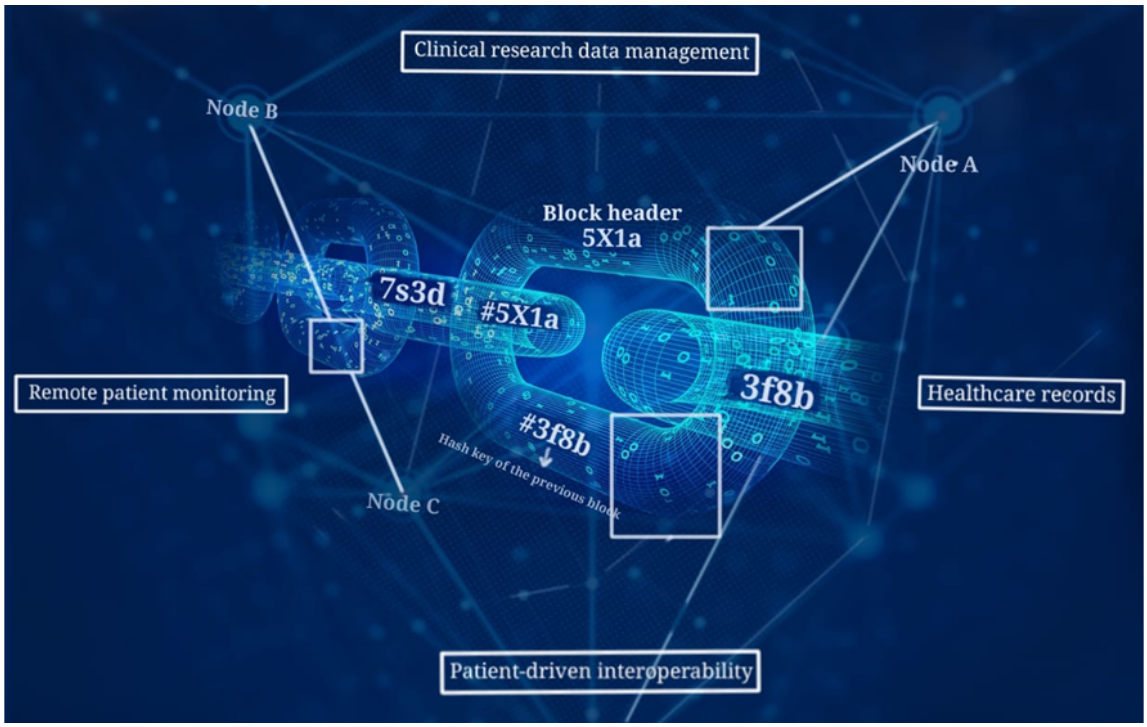
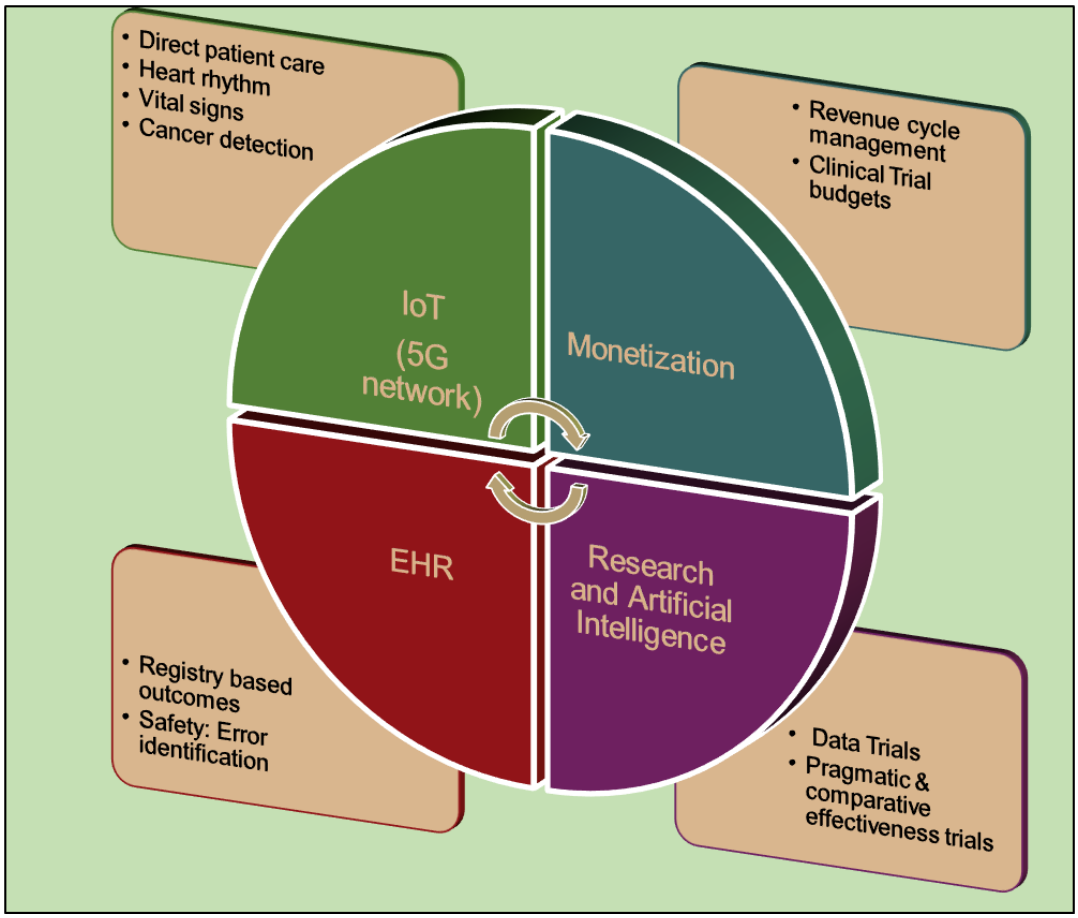


Figure 3. Areas of Blockchain utilization in healthcare.



The IoT Concept, Blockchain, and Clinical Trials

The integration of blockchains with other technologies such as artificial intelligence (AI), IoT, and big data management can be highly effective and act as a catalyst for innovation and increased efficiency, which is invaluable for the interpretation and management of data.

IoT refers to a world where people, things, and devices are all connected through the internet, allowing them to collect and exchange data seamlessly over the network. From coffee machines, wearable devices, and sensors to home security systems, IoT will likely change the way people live because 5G wireless technology (with high bandwidth and low latency) is becoming readily available. All the data streamed over the network can interlock without the need for human interactions. Future smart cities are based on IoT devices and applications. Additionally, it has been gaining considerable attention from stakeholders, investors, and various organizations owing to its unlimited application possibilities [19-21].

Data from sensors and wearable devices can revolutionize the way health care is viewed and delivered, especially in an era of patient-centered care, precision medicine, and individualized health care delivery. IoT can transform the approach to health care and take patient-centered care to a new level, where people can take charge of their health, providing patients and physicians with invaluable continuous real-time data about the physiological state and well-being of patients, ranging from data such as the heart rate, temperature, or sleeping habits to biochemical marker measurements in biofluids through various biosensor technologies, as well as sending alerts when certain thresholds are crossed or abnormal patterns are detected [22-27].

Using IoT technology in conjunction with blockchain technology can maximize its efficacy and potential uses. The massive amounts of various data streamed through different IoT devices can be used to collect large amounts of invaluable data for researchers to analyze and interpret. It could also be useful for data-hungry AI technology companies, public health surveillance, monitoring of disease outbreaks (eg, for monitoring COVID-19), epidemiology, and patient-oriented outcomes [28-33].

AI has been one of the key catalysts in health care innovation; for instance, researchers at the Massachusetts Institute of Technology made a ground-breaking discovery of a new antibiotic using AI technology through a trained deep learning model that was able to produce a powerful wide-spectrum antibiotic called Halicin [34]. However, one of the most critical shortcomings of AI and a crucial component for achieving revolutionary benefits is the requirement of tons of data for training its models to produce accurate and useful outputs. The combination of AI, IoT, and blockchain technologies can be powerful, where IoT devices provide the data (input), blockchains facilitate their transmission to various machine learning, and deep learning models can translate these data into extremely useful outputs. Some of the newer developments in machine learning are significantly driving blockchains to be better integrated with AI in the health care field. This enables

improvements in the security, privacy, functionality, and operational aspects of blockchain technology for health care applications [12,35,36].

Clinical Scenarios, Potential Applications, and Conclusions

If Chenoa had IoT devices that could monitor her vital signs (particularly temperature) 24/7 and the information was transmitted live via a blockchain to live monitors powered by machine learning algorithms, then her temperature fluctuations and trends would have prompted a return to the transplant center much sooner. Although research on nanotechnology and IoT-based sensors for detecting cancers (or relapses) is in its infancy, IoT devices for monitoring vital signs have already proved their effectiveness in monitoring patient physiology. Moreover, if Chenoa's EHR was available to all the treating clinicians across the country via a blockchain, it may have triggered a sense of urgency in the local physician treating her.

For Khaled, having his EHR not readily available to a treating emergency room physician is perhaps a classic example of a disjointed medical ecosystem. If there are IoT devices detecting potassium, oxygen, and vital signs in a heart failure patient, they could be instrumental in saving the lives of many patients with heart and kidney diseases. Moreover, having a living will and information on the power of attorney on a blockchain could be very helpful in certain end-of-life cases as well. Finally, for the hundreds of thousands of patients participating in clinical trials, informed consent on a blockchain could be very beneficial for the trial sponsors, ethics boards (eg, institutional review boards), and patient care providers.

Thus, the potential benefits of blockchains are limitless; however, concrete data on long-term clinical outcomes based on blockchains powered and supplemented by AI and IoT are yet to be achieved. Nonetheless, the implementation of blockchains as a novel way to integrate EHRs nationwide and manage common clinical problems in an algorithmic fashion has the potential of saving thousands of lives like those of Chenoa and Khaled.

Conflicts of Interest

None declared.

References

1. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. Decentralized Business Review. 2008. URL: http://bitcoin.org/bitcoin.pdf [accessed 2020-03-03]
2. Hoy MB. An introduction to the blockchain and its implications for libraries and medicine. Med Ref Serv Q 2017 Jul;36(3):273-279. [doi: 10.1080/02763869.2017.1332261] [Medline: 28714815]
3. Zhou T, Li X, Zhao H. Med-PPPHIS: blockchain-based personal healthcare information system for national physique monitoring and Scientific Exercise Guiding. J Med Syst 2019 Aug;43(9):305. [doi: 10.1007/s10916-019-1430-2] [Medline: 31410583]
4. Hylock RH, Zeng X. A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study. J Med Internet Res 2019 Aug;21(8):e13592 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
5. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating health Data sharing: a solution based on the Internet of Things and distributed ledger technologies. J Med Internet Res 2019 Jun;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
6. Zhuang Y, Sheets L, Shae Z, Tsai JJP, Shyu C. Applying blockchain technology for health information exchange and persistent monitoring for clinical trials. In: AMIA Annu Symp Proc. 2018 Presented at: AMIA Annual Symposium; November 3-7, 2018; San Jose, CA, United States p. 1167-1175.
7. Johnson M, Jones M, Shervey M, Dudley JT, Zimmerman N. Building a secure biomedical data sharing decentralized app (DApp): tutorial. J Med Internet Res 2019 Oct;21(10):e13601 [FREE Full text] [doi: 10.2196/13601] [Medline: 31647475]
8. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. JMIR Med Inform 2018 Dec;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
9. Satamraju KP, Malarkodi B. Proof of concept of scalable integration of Internet of Things and blockchain in healthcare. Sensors (Basel) 2020 Mar;20(5):1389 [FREE Full text] [doi: 10.3390/s20051389] [Medline: 32138380]
10. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. J Med Syst 2018 Jun;42(7):130. [doi: 10.1007/s10916-018-0982-x] [Medline: 29876661]
11. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
12. Kuo T, Gabriel R, Cidambi K, Ohno-Machado L. Expectation propagation logistic regression on permissioned blockchain (ExplorerChain): decentralized online healthcare/genomics predictive model learning. J Am Med Inform Assoc 2020 May;27(5):747-756 [FREE Full text] [doi: 10.1093/jamia/ocaa023] [Medline: 32364235]
13. Luna R, Rhine E, Myhra M, Sullivan R, Kruse CS. Cyber threats to health information systems: a systematic review. Technol Health Care 2016 Jan;24(1):1-9 [FREE Full text] [doi: 10.3233/THC-151102] [Medline: 26578272]

14. Argaw ST, Bempong N, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review. BMC Med Inform Decis Mak 2019 Jan 11;19(1):10 [FREE Full text] [doi: 10.1186/s12911-018-0724-5] [Medline: 30634962]
15. Tham I. Personal info of 1.5m SingHealth patients, including PM Lee, stolen in Singapore's worst cyber attack. The Straits Times. 2018. URL: https://www.straitstimes.com/singapore/personal-info-of-15m-singhealth-patients-including-pm-lee-stolen-in-singapores-most [accessed 2018-07-20]
16. Alam Khan F, Asif M, Ahmad A, Alharbi M, Aljuaid H. Blockchain technology, improvement suggestions, security challenges on smart grid and its application in healthcare for sustainable development. Sustain Cities Soc 2020 Apr;55:102018. [doi: 10.1016/j.scs.2020.102018]
17. Sultana M, Hossain A, Laila F, Taher KA, Islam MN. Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology. BMC Med Inform Decis Mak 2020 Oct;20(1):256 [FREE Full text] [doi: 10.1186/s12911-020-01275-y] [Medline: 33028318]
18. Boireau O. Securing the blockchain against hackers. Network Security 2018 Jan;2018(1):8-11. [doi: 10.1016/s1353-4858(18)30006-0]
19. Kamel BMN, Al-Shorbaji NM. On the Internet of Things, smart cities and the WHO healthy cities. Int J Health Geogr 2014 Mar;13:10 [FREE Full text] [doi: 10.1186/1476-072X-13-10] [Medline: 24669838]
20. Lynggaard P, Skouby KE. Complex IoT systems as enablers for smart homes in a smart city vision. Sensors (Basel) 2016 Nov;16(11):1840 [FREE Full text] [doi: 10.3390/s16111840] [Medline: 27827851]
21. Cirillo F, Wu F, Solmaz G, Kovacs E. Embracing the future Internet of Things. Sensors (Basel) 2019 Jan;19(2):351 [FREE Full text] [doi: 10.3390/s19020351] [Medline: 30654571]
22. Dimitrov DV. Medical Internet of Things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
23. Chen HS, Jarrell JT, Carpenter KA, Cohen DS, Huang X. Blockchain in healthcare: a patient-centered model. Biomed J Sci Tech Res 2019 Aug;20(3):15017-15022 [FREE Full text] [Medline: 31565696]
24. Kamel Boulos MN, Wilson JT, Clauson KA. Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. Int J Health Geogr 2018 Jul;17(1):25 [FREE Full text] [doi: 10.1186/s12942-018-0144-x] [Medline: 29973196]
25. Kang M, Park E, Cho BH, Lee K. Recent patient health monitoring platforms incorporating Internet of Things-enabled smart devices. Int Neurourol J 2018 Jul;22(Suppl 2):S76-S82 [FREE Full text] [doi: 10.5213/inj.1836144.072] [Medline: 30068069]
26. Yang Z, Zhou Q, Lei L, Zheng K, Xiang W. An IoT-cloud based wearable ECG monitoring system for smart healthcare. J Med Syst 2016 Dec;40(12):286. [doi: 10.1007/s10916-016-0644-9] [Medline: 27796840]
27. Coulby G, Clear A, Jones O, Young F, Stuart S, Godfrey A. Towards remote healthcare monitoring using accessible IoT technology: state-of-the-art, insights and experimental design. Biomed Eng Online 2020 Oct;19(1):80 [FREE Full text] [doi: 10.1186/s12938-020-00825-9] [Medline: 33126878]
28. Bhattacharya S, Singh A, Hossain MM. Strengthening public health surveillance through blockchain technology. AIMS Public Health 2019 Sep;6(3):326-333 [FREE Full text] [doi: 10.3934/publichealth.2019.3.326] [Medline: 31637281]
29. Singh RP, Javaid M, Haleem A, Suman R. Internet of things (IoT) applications to fight against COVID-19 pandemic. Diabetes Metab Syndr 2020 Jul;14(4):521-524 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.041] [Medline: 32388333]
30. Panarello A, Tapas N, Merlino G, Longo F, Puliafito A. Blockchain and IoT integration: a systematic survey. Sensors (Basel) 2018 Aug;18(8):2575 [FREE Full text] [doi: 10.3390/s18082575] [Medline: 30082633]
31. Dwivedi AD, Srivastava G, Dhar S, Singh R. A decentralized privacy-preserving healthcare blockchain for IoT. Sensors (Basel) 2019 Jan;19(2):326 [FREE Full text] [doi: 10.3390/s19020326] [Medline: 30650612]
32. Madanian S, Parry D. IoT, cloud computing and big data: integrated framework for healthcare in disasters. Stud Health Technol Inform 2019 Aug;264:998-1002. [doi: 10.3233/SHTI190374] [Medline: 31438074]
33. Catherwood PA, Steele D, Little M, Mccomb S, Mclaughlin J. A community-based IoT personalized wireless healthcare solution trial. IEEE J Transl Eng Health Med 2018 May;6:1-13 [FREE Full text] [doi: 10.1109/JTEHM.2018.2822302] [Medline: 29888145]
34. Stokes J, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep Learning approach to antibiotic discovery. Cell 2020 Feb;180(4):688-702.e13 [FREE Full text] [doi: 10.1016/j.cell.2020.01.021] [Medline: 32084340]
35. Chen X, Ji J, Luo C, Liao W, Li P. When machine learning meets blockchain: decentralized, privacy-preserving secure design. United States: IEEE; 2018 Presented at: IEEE International Conference on Big Data (Big Data); Decembe 10-13, 2018; Seattle, WA, United States p. 1178-1187. [doi: 10.1109/bigdata.2018.8622598]
36. Li Z, Liu J, Hao J, Wang H, Xian M. CrowdSFL: a secure crowd computing framework based on blockchain and federated learning. Electronics 2020 May;9(5):773. [doi: 10.3390/electronics9050773]

Abbreviations

AI: artificial intelligence
EHR: electronic health record

HSMs: hardware security modules
IoT: Internet of Things

Edited by G Eysenbach; submitted 04.05.20; peer-reviewed by M Beg, C Reis, TT Kuo; comments to author 26.10.20; revised version received 20.01.21; accepted 03.04.21; published 02.11.21.

Please cite as:
Fatoum H, Hanna S, Halamka JD, Sicker DC, Spangenberg P, Hashmi SK
Blockchain Integration With Digital Technology and the Future of Health Care Ecosystems: Systematic Review
J Med Internet Res 2021;23(11):e19846
URL: https://www.jmir.org/2021/11/e19846
doi:10.2196/19846
PMID:34726603

Original Paper

Mapping Information Needs of Patients With Sexually Transmitted Infections Using Web-Based Data Sources: Grounded Theory Investigation

Pavankumar Mulgund[1], PhD; Raj Sharman[1], PhD; Sandeep Purao[2], PhD; Sagarika Suresh Thimmanayakanapalya[1], MS; Peter Winkelstein[3], MD, MBA

[1]Department of Management Science and Systems, State University of New York at Buffalo, Buffalo, NY, United States

[2]Department of Information & Process Management, Bentley University, Waltham, MA, United States

[3]Department of Pediatrics, Jacobs School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY, United States

Corresponding Author:
Pavankumar Mulgund, PhD
Department of Management Science and Systems
State University of New York at Buffalo
325 Jacobs Management Center
Buffalo, NY, 14260
United States
Phone: 1 7166453271
Email: pmulgund@buffalo.edu

Abstract

Background: According to the World health organization (WHO), more than 1 million sexually transmitted infections (STIs) are acquired each day across the world. The incidence rates of STIs in the United States are at a record high for the fourth consecutive year. Owing to the stigma associated with the incidence of STI, there is a general reluctance to seek information in person. Instead, web-based information sources remain the primary avenues of information-seeking. However, these sources are designed without a comprehensive understanding of the information needs of individuals who have contracted STIs.

Objective: This study aimed to investigate the information needs of individuals who have or suspect they have contracted an STI. A better understanding of their information needs can drive the design of more effective digital interventions.

Methods: This is a qualitative and analytical study of 549 transcripts (consisting of queries posted over the last 10 years) from web-based forums of the American Sexual Health Association (ASHA), which allows patients, volunteers, and health care providers connect anonymously. The analysis follows a grounded theory (GT) approach with multiple coding stages to uncover categories and themes.

Results: Three categories of information needs emerged. The first two, *clinical* and *logistical*, are similar to other contexts. However, our analysis shows that there is a significant need for the last category—*psychosocial* information. Approximately 59% of instances are linked to concerns such as confusion, discretion, remorse, and others. These needs vary across the stages of a patient's journey from symptom manifestation to treatment maintenance.

Conclusions: Responding to the needs of individuals who have or suspect they have contracted an STI requires compassionate and personalized responses (beyond factual clinical and logistical information). Web-based forums provide anonymity but do not adequately incorporate mechanisms, practices, or incentives to respond to diverse psychosocial concerns. Innovative approaches to add such support can make the digital interventions more effective for this group of individuals.

(J Med Internet Res 2021;23(11):e30125) doi:10.2196/30125

KEYWORDS

information needs; sexually transmitted diseases; patient journey maps; health information seeking; stigmatizing disorders; online forum; sexually transmitted infection; American Sexual Health Association; grounded theory; stigma

Introduction

Background

The incidence of sexually transmitted infections (STIs) remains high across the globe and within the United States [1,2]. According to a report released in January 2021, the US Centers for Disease Control and Prevention estimated that on any given day in 2018, one in 5 people in the United States had an STI [3]. Consequences include acquiring HIV, mother-to-child transmission resulting in stillbirth, congenital deformities, and other complications such as cervical cancer [4]. Within the United States, the annual cost of treating and managing STIs is estimated to be more than US $16 billion [5].

Despite this scale and such severe consequences, many individuals who have or suspect they have contracted an STI do not readily consult their physician. Owing to the stigma associated with the incidence of STIs, their primary source of information appears to be web-based forums, where they can anonymously seek information from other patients, volunteers, and health care providers [6]. These forums facilitate questions and answers while protecting the identity of individuals. However, individual contributors and designers of these forums do not appear to possess an understanding of the information needs [7] of this population.

In this paper, we developed an in-depth understanding of the information needs of individuals who have or suspect they have contracted an STI following a grounded theory (GT) approach. The analysis relies on secondary data (queries posted) on a web-based forum. The results are described as categories of information needs associated with stages of the patient journey. The results are important because they inform our understanding of how individuals search for information about taboo-ridden disorders; and they are useful because they can inform the design of digital interventions in response.

Rationale

Information-seeking describes an individual's conscious effort to acquire information in response to a need or gap [8]. An information need emerges from the realization that one's knowledge is insufficient to satisfy a goal at hand and then culminates in locating information that contributes to understanding and meaning [9]. Information-seeking is a complex phenomenon characterized by the interplay of seekers' personal needs, learning styles, available resources, and affective components, among others [10]. It begins when one realizes the need for information and ends when that need is believed to have been satisfied [11].

Health information–seeking has been defined as a "problem-focused coping strategy … adopted by individuals as a response to a threatening situation" [12], which may also include "the urge to confront oneself with the threatening situation by means of seeking more information about it" [13]. The problem is important because health care providers as well as researchers are interested in understanding "how and why individuals obtain health information, where they go to retrieve such information, what particular types of information they prefer, and how the health information sought is used" [14].

Much prior scholarship points out that information needs can vary on the basis of the disease or condition. In addition to clinical information like symptoms and treatments; individuals may also seek emotional support, anonymity, and privacy, especially for conditions (such as STIs) whose incidence invokes stigma [15,16]. In these cases, individuals answering the questions (physicians) may also prefer anonymity. Researchers have attempted several innovative techniques to encourage a more open dialogue with patients with STIs about their information needs. In addition to the traditional methods such as focus groups and structured interviews, novel techniques such as concept cards [17], photo-based projective techniques [18], computer-assisted video elicitation [19], and vignettes [20] have been used. However, the success of these methods is associated with the facilitator's skills and can generate incomplete or incorrect outcomes [21]. Consequently, eliciting and understanding the information needs of a patient with an STI remains a challenging endeavor. This is the precise challenge we undertake in this study.

Methods

Methods Overview

We used a GT approach relying on secondary data analysis [22]. Use of secondary data sources in GT approaches is quite common. Following the idea that all is data, grounded theorists have used a variety of sources such as documents [23], text, and transcripts of written material [24,25]. However, 2 issues assume importance when conducting a GT approach using secondary data. First, the secondary data source should be rich in terms of volume and diversity to develop a generalizable theoretical framework. Second, the researchers should use contextual positioning [26] to collect the data for analysis, which involves approaching data reflectively to establish the context. In this study, we address both these issues, as described in the following sections.

We adopted a GT approach because it is appropriate for deciphering shades of meaning, implied pointers, and included efforts to traverse between an utterance or a phrase or sentence to the larger context of the discourse. We considered the possibility of automated analyses, which offers scale and efficiency, but decided not to pursue it because, in its current level of maturity, automated analysis may not produce as conceptually rich observations as GT. Such methods also require significantly larger data sets and may lead to spurious correlations at lower volumes of data.

As a descriptive approach that emphasizes the discovery of central concepts, GT is well suited to uncover the information needs of patients with STIs. GT is also a good fit because of a lack of prior theoretical work that explains the information needs of individuals who are affected by a taboo-ridden condition such as an STI. Although much prior work [27,28] has discussed patients' information needs at a more general level, it has not accounted for the unique challenges that the stigmatizing disorders (such as STIs) present. When applied rigorously, GT is a robust methodology that fits well in practical situations with a real-world orientation. It is an iterative approach, where researchers interlace data collection and analysis until

"concepts" emerge. The concepts are interpreted in response to the research goal, which in this case is to understand the information needs of individuals who have or suspect they have contracted an STI.

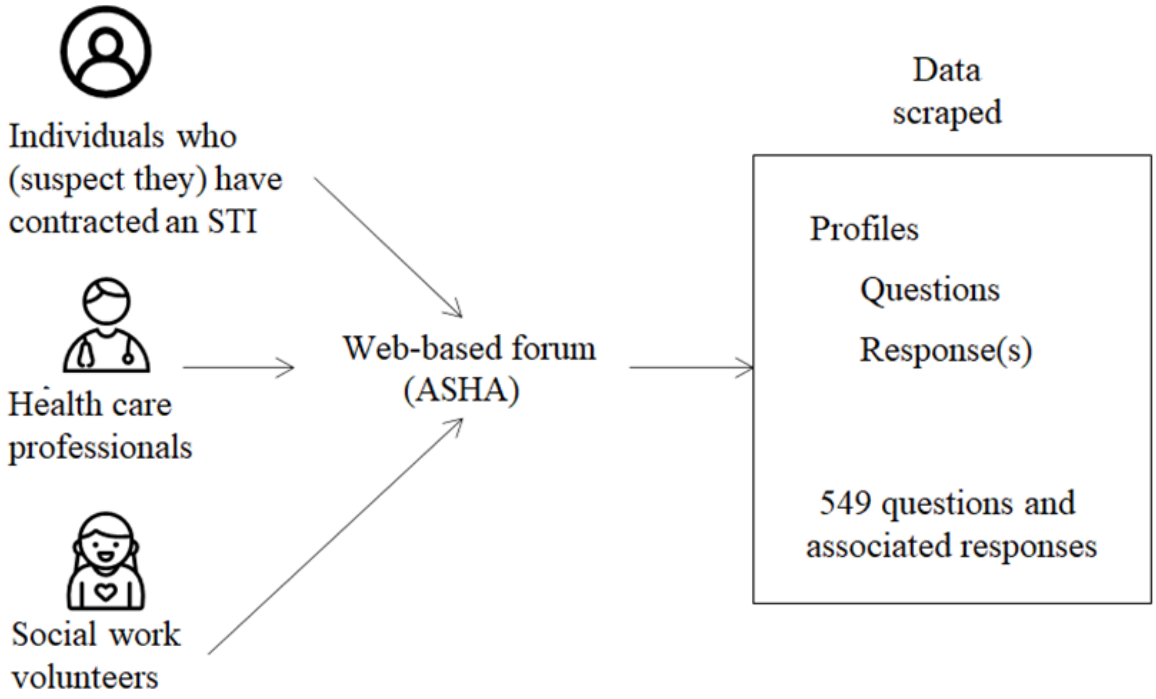
As patterns begin to emerge, the concepts are classified and tied to the context, which in this case is patient journeys, grounded in the realities of seeking information when faced with different uncertainties. We applied the GT approach with the following specific decisions.

Data Collection

We collected transcripts of questions and responses from the Question-and-Answer forum of the American Sexual Health Association (ASHA), a nonprofit organization operating in the United States. In line with its mission to "empower individuals, families, and communities to achieve sexually healthy lives through education and advocacy," ASHA offers a web-based community that connects patients, families, friends, and caregivers to support individuals with STIs. The web-based portal sponsored by ASHA and maintained by the Inspire community allows individuals who have or suspect they have contracted an STI to ask STI-related questions. The volunteers who respond to questions include caregivers (nurses and physicians), social workers, health communication experts, and individuals with prior experience managing their STIs. Although the content can be generated by any member participating in the community, the threads are moderated by experts. The portal requests nonidentifiable information from the contributors, such as age, gender, and location.

After cleansing the data, we obtained 549 transcripts, which comprised questions and associated responses. For each transcript we collected (1) the query posted by the individual who has or suspects he/she has contracted an STI, (2) responses to the query posted by other participants on the portal, and (3) nonidentifiable information of each contributor on their profile (Figure 1).

Figure 1. Data collection from the web-based forum. ASHA: American Sexual Health Association, STI: sexually transmitted infection.



Data Analysis

Data analysis included generating summaries of participant demographics, followed by a characterization of questions for each demographic, and finally, an analysis of the content of the transcripts. During this content analysis, the first phase involved analyzing the complete set (549 transcripts) sequentially with open coding (extraction of keywords and phrases, and assigning codes to each [29]), which resulted in 115 unique codes. The results were examined to identify a subset (73 transcripts) on the basis of the uniqueness of codes they generated. This selection was thus nonprobabilistic, and followed theoretical sampling guidelines stipulated by prior research [30]. The second phase of the analysis (with 73 transcripts) involved revisiting these transcripts for more intense analysis with the "constant comparative method" [29]. We revisited the transcripts to compare codes extracted from the initial transcripts against those from later transcripts. This phase also involved axial coding (grouping the codes to identify code clusters [29]), which resulted in 43 categories. The last phase of the analysis process involved selective coding, which included identifying and articulating higher-level themes [29]. Theoretical saturation was observed in this iterative process as the number of new codes dropped to near-zero. Across these phases, we generated several memos and representations to record details, and capture different interpretations. The effort was managed with the Atlas.ti software. Figure 2 outlines the data analysis process.

Figure 2. Data analysis across 3 phases.

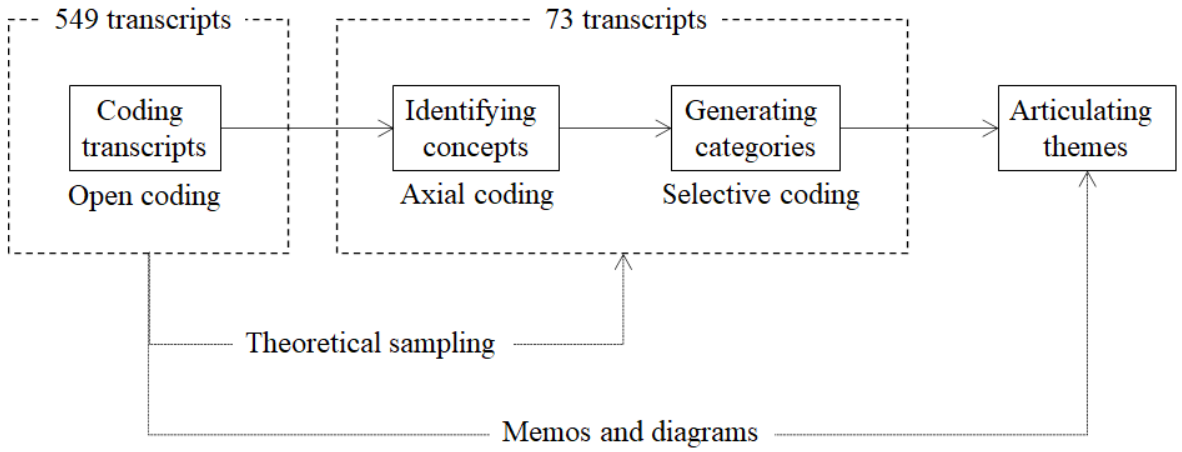


Table 1 shows an example of outcomes from the coding process to illustrate the path from raw data to the discovery of concepts. The effort was data-driven but influenced by the authors' exposure to prior research [31].

An important ancillary outcome from our analysis was the recognition that the individual patients appeared to follow a journey that started with symptom manifestation and progressed through several stages to ongoing maintenance of the disease.

Table 1. An example of data coding.

Raw data	Phase 1: open coding	Phase 2: axial coding	Phase 3: selective coding
"I just found out today that I tested positive and I am lost for words I can't stop crying I don't know what to do I've been married for 3 years made the biggest mistake of my life went out my marriage now I don't know what to do or how to tell my partner"	• Feeling miserable • Feeling anxious • Feeling lost • Feeling remorseful	• Managing negative sentiments	Psychosocial information needs
I'm driving myself crazy on google."	• Struggling with lack of relevant information • Overwhelmed by internet-based content	• Incomplete information • Concern about misinformation	
"The thought of even meeting a man and having to explain it all over again to a new person never feels good."	• Fear of disclosing information • Worrying about not being able to have a child	• Worries about future relationships • Anxiety about bleak prospects	

Results

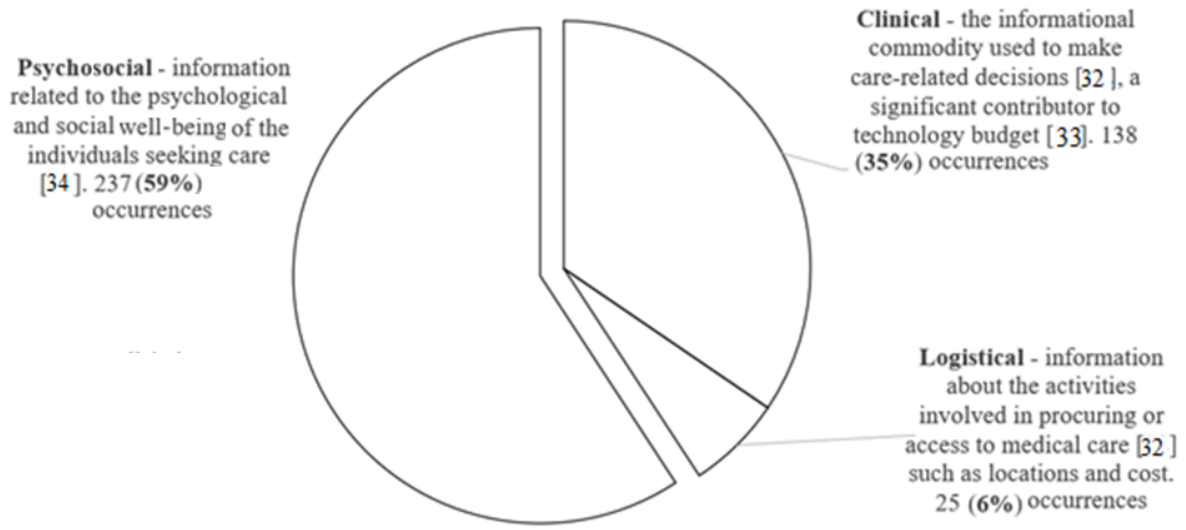
Participant Characteristics

The participants in the forum, 87 individuals (47 women, 29 men, and 11 not specified) ranged in age from 18 to 74 years (women: 20 to 52 years, median ~29 years, average ~30 years, mode ~34 years; men: 18 to 74 years, median ~43 years, average ~42 years, mode ~49 years). As many as 2 out of 3 queries were posed by individuals aged between 18 and 35 years. Only 5 individuals described themselves as part of the LGBT (lesbian, gay, bisexual, and transgender) community (self-reported, does not necessarily suggest lack of participation from the most vulnerable population segment). The data do not include individuals who may have visited the forum to read posts from others (without asking or responding to a question). Despite these caveats, 62% of the posted questions were from female information seekers. Most frequent infections discussed were herpes, syphilis, chlamydia, genital warts, and gonorrhea (a total of 503 occurrences). Very few (3 occurrences) of preventive behaviors discussed and included condom use and vaccinations.

Categories of Information Needs

The key finding was the different categories of information needs. Here, the unexpected finding was that as many as 59% of the queries belonged to the category of *psychosocial information needs*, indicating a set of diverse concerns. Figure 3 summarizes these findings.

Figure 3. Categories of information needs.



Category 1: Clinical Information Needs

The first category, *clinical information needs* is defined as the informational commodity used to make care-related decisions [32] and accounts for a significant part of information technology budgets within the health care organizations [33]. It included details about the conditions, symptoms, risk factors, epidemiology (how the disease spreads), possible differential diagnosis, clinical presentation, and prognosis. The following excerpt illustrates this further.

> *I had unprotected sex with a man last week and afterwards, he told me that his previous partner has HSV-2. He told me that she hasn't had an outbreak in years and that he's never had an outbreak. Should I be concerned about contracting HSV-2? I was going to get tested in a month and see what the results are. Any information will help.*

The clinical information sought may include diagnosis and initial treatment and other relevant information such as treatment costs and access. The following excerpt illustrates this further.

> *If I've catched chlamydia conjunctivitis only on my left eye as a first contact with the germ, for example from touching with my hand a infected fluids and then with the same hand touching my eye (No penetration, or oral sex). Is it possible the chlamydia goes to my sexual organs with the pass of time? (Considering that the first contact with the germ was in the eye).*

The individual seeking such information often shared personal circumstances in an effort to seek specific clinical information and even offered advice to others. The following excerpt illustrates this further.

> *Today I begin the transition to a new apartment as my disease has progressed to the point of not being able to climb stairs anymore. I can barely walk on level ground as the bacteria that has been living in my body for apparently many, many years has irreversibly damaged my nervous system. If you have been sexually active in the past 20 years, please get tested for Syphilis before it gets to the "tertiary" stage and you wind up disabled like me.*

Category 2: Logistical Information Needs

The second category, *logistical information needs* is defined as information about activities involved in procuring and access to medical care [32]. It included 2 broad subcategories. The first was *financial considerations* such as insurance coverage or out-of-the-pocket cost of medication and treatment. The following excerpt illustrates this further.

> *I've been advised by 3 doctors that I NEED surgery for the removal of anal warts and most likely are internal, anyone been through the surgery in the anal area? What's the cost like? process? HELP ME I am so scared, I've been infected since December of 2010 I've had a few freezing treatments, it never worked for me, now in need of surgery.*

The second subcategory was information about providers, testing facilities, appointment timelines, and reviews about providers and test efficacy. Questions about educational resources, contraceptives, and helplines were also discussed. The following excerpt illustrates this further.

> *I am unable to find information on any website with a suggested timetable from exposure for STD testing. Does anyone know this information/suggestion where to look? Which diseases should be retested in another period of time?*

Category 3: Psychosocial Information Needs

The third category, *psychosocial information needs* is defined as the information related to the psychological and social well-being of the individuals seeking care [34]. These needs were not only the most complex but also the most idiosyncratic [35,36]; they are also highly sought after and least well-addressed, as evidenced by the appearance of high-sentiment words such as "shocked," "worried," "depressed," "scared," and "fear" in 82% of the transcripts. Psychosocial needs emerge to cope with the negative emotions involved with a general feeling of not being healthy [37]. Such

needs are amplified in stigmatizing topics such as those relating to STIs. “Psychosocial needs” vary considerably through each phase of the patient's journey. Several excerpts illustrate this further.

> *I had unprotected sex 3 days ago and suddenly have a tingling/pain feeling in my inner thig/groin and kind of throughout my leg. I am kind of freaking out because I'm thinking the worse, i might just be paranoid can you please help me figure out what this is. No other symptoms just what i described above.*

The concerns expressed by participants included the shock of becoming infected with STIs, the despair and anger associated with the outcome.

> *I feel like my world is shattered. I’ve cried an ocean and am still trying to process all this. I’m just looking for a support group on how to cope with this*

Participants described the feeling of being dirty or contaminated with extreme negative emotions, including suicidal tendencies, as highlighted in the transcripts below.

> *I just started dating another woman about 2 months ago when I noticed a small bump... like a pimple... on my penis. I went to my doctor thinking it was a benign skin tag. I had skin tags removed from other parts of my body in years past. I was shocked... absolutely shocked when the doc told me that the bump was a genital wart. I feel dirty, trashy and disgusted.*

Some worried about being judged by the providers and talked about embarrassments because of disclosing unpleasant and discrediting details about prior sexual behavior. The following transcripts illustrate these issues.

> *I have seen multiple doctors and most have said they look like very tiny warts (almost microscopic) but are too small to treat since any of the treatments used would ruin healthy surrounding skin. Two docs even said not worry since “I look like a healthy person” and they should clear on their own. That was 15 years ago. Some docs think its from another type of genital hpv other than 6 or 11. This is so frustrating and I am now in my 40s. I have seen over 8 doctors.*

Two other issues were also frequently discussed. These included worries about treatment adherence with complex treatment regimens and fear of dealing with associated pain.

> *I know it’s not the end of the world and it’ll get better, but it’s just fresh and the pain in HORRIBLE, anyone have advice?? Or remedies to make the outbreaks less pain.*

Concerns about revealing to current and prospective partners is another common concern. Although there are regulations for partner communication and contact-tracing, there is inadequate guidance about on how to (1) engage in such difficult conversation and (2) deal with the anxiety of revealing such information to a partners.

These concerns are especially pronounced among middle-aged and senior information seekers.

> *The thought of even meeting a man and having to explain it all over again to a new person never feels good. The thought of being single after not being single for so long is anxiety-provoking enough.*

There was also worry about postincidence sex life and the fear of passing the infection to others. This was amplified among participants who feared the negative consequences to their reproductive health.

> *I was hoping to find other moms who have breastfeed their children successfully with HSV-2. I have been exclusively pumping and bottle feeding due to an article I read about contracting HSV through breastfeeding and so I’ve stopped but desperately want to breastfeed and was hoping I’m being completely irrational about this. I’ve never had an outbreak on my nipples nor did I realize that was possible. Hoping to be able to breastfeed again without risking my baby’s health.*

Patient Journey

The ancillary finding from the analysis—the patient journey—also revealed how individuals in this group progressed through different stages. We observed several references about patient stages (temporal progression of events, feelings, and things) in the question transcripts during the coding and memoing phases. On mapping the emergent codes to patient stages, we arrived at a tentative flow of patients seeking information on STIs. Subsequent literature review revealed the concept of patient journey maps and its fit with the goal of this study. Therefore, drawing on ideas from patient journey maps [38], we developed a conceptualization of how individuals progress through different stages of experience [39-41].

Adopting a clinical focus, prior scholarship has also created disease progression flows highlighting changes in signs and symptoms over time [42]. Although there is some alignment between patient journey and disease progression, patient journey map is a broader concept and presents the various stages of patient experience, including touchpoints with care providers, and feelings and emotions they experience. We present the details below.

The journey begins when an individual experiences symptoms, which may lead to evaluative tests and diagnosis, and an exploration of treatment options before reaching a treatment decision. Follow-ups for any troubleshooting would eventually lead ongoing maintenance of the disease. These stages were identified by examining the codes and synthesizing the ideas across transcripts. Figure 4 summarizes the patient journey.

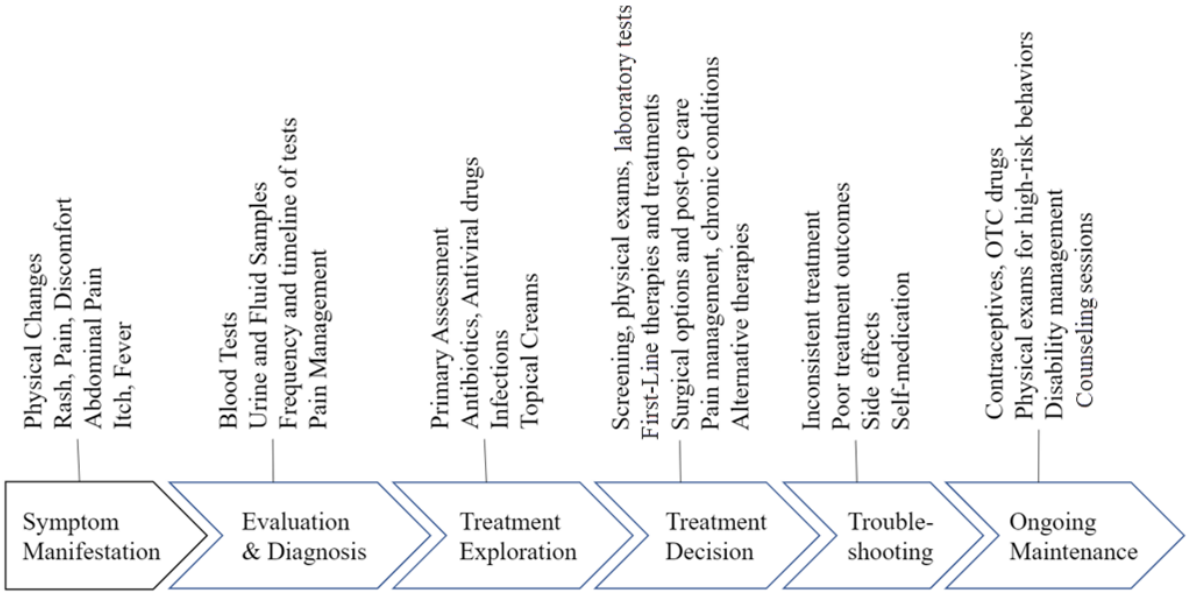
Figure 4. Patient journey.

The journey map concept has been used in prior research to explore needs across different stages in case of patients with cancer [43,44] or patients undergoing surgical interventions [45]. With the help of these ideas, we were able to interpret our findings; for example, by mapping the types of questions against different stages of the patient journey (Figure 3). Specifically, we found that the clinical information and psychosocial information needs varied considerably across the stages of the patient's journey map, whereas logistical information needs did not.

Information Needs Across the Patient Journey

The specific types of information needs for each category varied across the stages of the patient journey. For example, clinical information needs in the evaluation and diagnosis stage were about tests, samples, and pain management; these changed to treatment outcomes and side-effects during the troubleshooting stage. Figure 5 summarizes these needs.

Figure 5. Clinical information needs during the patient journey. OTC: over-the-counter.



The psychosocial information needs during the evaluation and diagnosis stage were about loss of self-control, seeking cause and blaming, and feeling contaminated; these transitioned to worries about treatment adherence, poor self-image, and disclosing information to partners during the troubleshooting phase. Figure 6 summarizes these needs.

The last category of information needs (logistical information) was not analyzed in this manner because of the small number of instances. Relative frequencies of the two categories (clinical and psychosocial information needs) were computed for each stage. Table 2 summarizes this information.

Figure 6. Psychosocial information needs during the patient journey. STI: sexually transmitted infection.

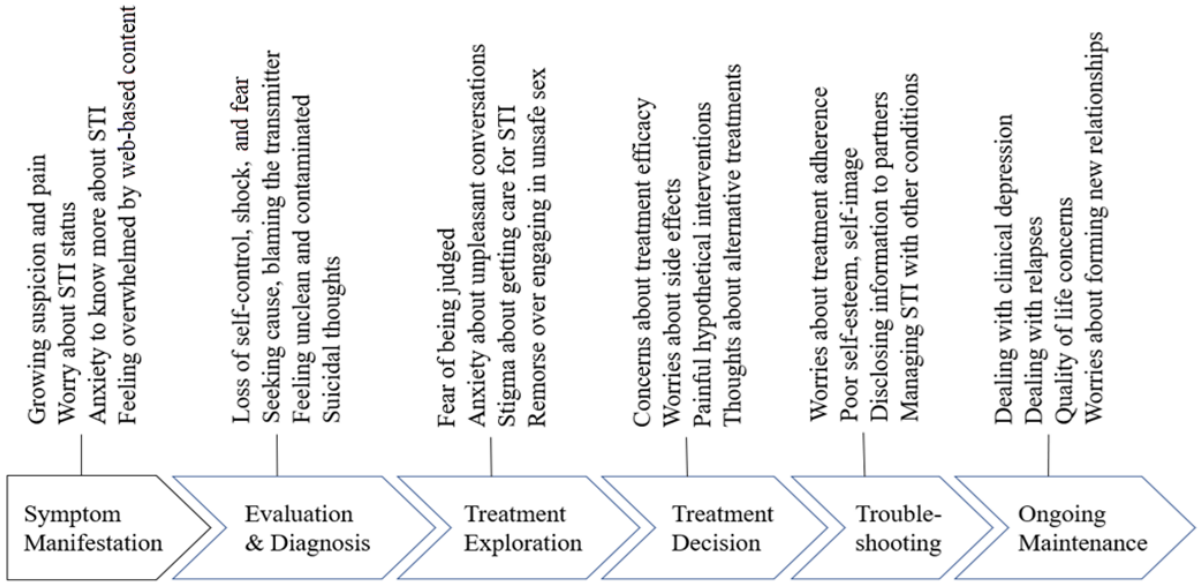


Table 2. Relative frequencies of information needs across stages of the patient journey.

Journey stage	Symptom manifestation, %	Evaluation and diagnosis, %	Initial treatment exploration, %	Treatment decision, %	Troubleshooting, %	Ongoing maintenance, %
Clinical needs	36.22	16.69	21.86	8.32	12.66	4.23
Psychosocial needs	31.50	18.32	12.46	11.44	19.20	7.12

Discussion

Principal Findings

This research adds to a stream of work that has explored information needs and information-seeking in general [46,47] and for specific populations [48]. Some of these studies have explored the information needs of women with HPV [49], health information needs of ethnically diverse adolescent patients with STIs [50], and preventive information needs among HIV-negative individuals [51]. Our work is unique in that it provides a comprehensive depiction of the information needs and includes three categories, *clinical*, *logistical*, and *psychosocial*, which are mapped across each stage of the patient journey.

A key finding is a complex set of psychosocial information needs, some of which vary with each stage of the patient journey and others that persist across stages. More importantly, we found that psychosocial information needs are most critical for individuals who have or suspect they have contracted an STI. This category of information needs remains least explored in prior work and is likely to present significant barriers to information access, further exacerbated by patients' general reluctance to seek in-person help and the lack of information sources that cater to these needs. Unlike clinical and logistical information needs that are well catered to by current information sources, there appears to be a lack of appropriate or tailored sources that can respond to the psychosocial information needs. Additionally, the study also points to a lack of emphasis on information-seeking for preventive behaviors.

Implications for Practice

This work has significant implications for practice by informing the design and development of information sources. Current information sources (general and specialized search engines and information portals) can respond to the patients' clinical and logistical information needs but do not consider their psychosocial information needs. Our findings highlight their relative importance of this category of information needs (approximately 59%) from a patient's perspective and can therefore encourage future designers to develop information sources as well as discussion forums that respond to this deficiency. Our findings are timely as we witness a surge in the number of people seeking information on the internet.

From an information-seeking perspective, our study underscores the need to cater to the emotional challenges of the individuals who have or suspect they have contracted an STI. Broadly, our study alludes to the need to understand and explore the information needs of individuals struggling with stigmatizing conditions. More research is warranted to explore similarities and differences among such disorders, and how we can respond to the information needs from these populations.

Limitations

There are limitations to this work. First, the study relies on data gathered from ASHA—one of the largest forums for patients with STIs in the United States. Even though ASHA contains a large volume of questions and responses, it may not reflect all of the information needs of patients with STIs. Analyzing information from other websites may provide additional insights. Further, ASHA caters mostly to people living in the United

States. Therefore, the results may not entirely reflect the differences in the information needs of individuals in other settings.

Second, although the study uncovered the *information needs* for individuals in this group, it cannot fully describe their *information-seeking behaviors*. Third, this work consolidates the patient journey into a single representation. Although it is grounded in rigorous and systematic mapping, the patient journey map is a generalized pathway, and individual patients may not always conform. The findings of this study may not apply to such nonconforming patients. Furthermore, this work does not account for variation among different STIs. Future work can examine additional sources, explore information-seeking behaviors, and capture pathways for subgroups of patients. Lastly, although it is unlikely that information portals (such as WebMD) instead of and web-based forums (such as ASHA) can address all the information needs of individuals from this group [50], it is important to explore how individuals may combine these sources.

Conclusions

The findings of this study emphasize that the information needs of individuals who have or suspect they have contracted an STI includes a range of information, including logistical and psychosocial queries, and not just clinical information. Psychosocial information, which is diverse and idiosyncratic, is sought most frequently. These information needs vary across the patient journey stages. The findings have significant practical implications for organizations providing web-based medical information and designing forums that allow patients, volunteers, and providers to connect. The incidence of an STI represents an example of an outset of a stigmatizing medical condition that prevents open information-seeking. This work highlights several specific components of information needs and some peculiar problems that are associated with such stigma. Based on the classification of information needs and its mapping to the patient journey, future work can contribute to the design of more compassionate and personalized responses to information needs with emerging technologies such as conversational AI and others.

Authors' Contributions

All authors have made sufficient and meaningful contributions to the conception, design, data collection and analysis, and drafting and editing the manuscript. All authors have reviewed the final version and consent to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Sexually Transmitted Infections (STIs). World Health Organization. 2019 Jun 14. URL: http://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis) [accessed 2021-01-06]
2. Americans Are Getting STDs at Record Rates. Time. 2018 Aug 28. URL: https://time.com/5379165/std-rates-sex/ [accessed 2021-01-06]
3. CDC estimates 1 in 5 people in the U.S. have a sexually transmitted infection. Centers for Disease Control and Prevention. 2021 Jan 25. URL: http://www.cdc.gov/media/releases/2021/p0125-sexualy-transmitted-infection.html [accessed 2021-04-09]
4. Casper C, Crane H, Menon M, Money D. HIV/AIDS Comorbidities: Impact on Cancer, Noncommunicable Diseases, and Reproductive Health. In: Holmes KK, Bertozzi S, Bloom BR, Jha P, editors. Major Infectious Diseases (3rd edition). Washington, DC: The International Bank for Reconstruction and Development / The World Bank; 2017.
5. Incidence, Prevalence, and Cost of Sexually Transmitted Infections in the United States. Centers for Disease Control and Prevention, National Center for HIV Viral Hepatitis STD and TB Prevention. 2013. URL: https://npin.cdc.gov/publication/incidence-prevalence-and-cost-sexually-transmitted-infections-united-states [accessed 2020-12-13]
6. Gray N, Klein J, Noyce P, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
7. Alzougool B, Chang S, Gray K. Towards a comprehensive understanding of health information needs. Electron J Health Inform 2008;3(2):15 [FREE Full text]
8. Miller VD, Jablin FM. Information seeking during organizational entry: influences, tactics, and a model of the process. AMR 1991 Jan;16(1):92-120. [doi: 10.5465/amr.1991.4278997]
9. Leu D, Kinzer C, Coiro J. Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In: Theoretical Models and Processes of Reading (5th edition). Newark, DE: International Reading Association; 2004:1570-1613.
10. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Libr Assoc 2005 Jul;93(3):353-362 [FREE Full text] [Medline: 16059425]
11. Krikelas J. Information-seeking behavior: Patterns and concepts. Drexel Libr Q 1983;19(2):5-20.
12. Rees CE, Bath PA. Meeting the information needs of adult daughters of women with early breast cancer. Patients and health care professionals as information providers. Cancer Nurs 2000 Feb;23(1):71-79. [doi: 10.1097/00002820-200002000-00012] [Medline: 10673810]

13. Rees CE, Bath PA. Information-seeking behaviors of women with breast cancer. Oncol Nurs Forum 2001 Jun;28(5):899-907. [Medline: 11421149]
14. Lambert S, Loiselle C. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
15. Allan K, Burridge K. Forbidden Words: Taboo and the Censoring of Language. In: Cambridge University Press. Cambridge: Cambridge University Press; 2006.
16. MacAndrew R, Martínez R. Taboos and Issues. Chicago, IL: LTP; 2001.
17. Sutton B. Playful cards, serious talk: a qualitative research technique to elicit women's embodied experiences. Qualitative Research 2011 Apr 14;11(2):177-196. [doi: 10.1177/1468794110394070]
18. Caleb-Solly P, Flind A, Vargheese J. Cameras as cultural probes in requirements gathering? Exploring their potential in supporting the design of assistive technology. 2011 Presented at: Inth International Symposium on Computer-Based Medical Systems (CBMS); June 27-30, 2011; Bristol p. 2011-2024. [doi: 10.1109/cbms.2011.5999158]
19. Spezialetti M. Putting People in the Picture: Building Requirements Gathering, Design SpecificationCommunication Skills with Video Scenarios. 2021 Presented at: SIGCSE '21: The 52nd ACM Technical Symposium on Computer Science Education; March 13-20, 2021; Virtual URL: https://doi.org/10.1145/3408877.3432553 [doi: 10.1145/3408877.3432553]
20. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
21. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019;7(2):e000057 [FREE Full text] [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
22. Heaton J. Secondary analysis of qualitative data. SAGE Research Methods Foundations 2019. [doi: 10.4135/9781526421036849600]
23. Glaser B. Basics of Grounded Theory Analysis: Emergence vs. Forcing. Mill Valley, CA: Sociology Press; 1992.
24. Bernard H, Ryan G. Qualitative and Quantitative Methods of Text Analysis. In: Hand-book of Method in Cultural Anthropology. Lanham, MD: Altamira Press; 1998:595-646.
25. Birks M, Mills J. Grounded Theory: A Practical Guide (2nd edition). Thousand Oaks, CA: Sage Publications; 2015.
26. Ralph N, Birks M, Chapman Y. Contextual positioning. SAGE Open 2014 Sep 29;4(3):215824401455242. [doi: 10.1177/2158244014552425]
27. Tang PC, Newcomb C, Gorden S, Kreider N. Meeting the information needs of patients: results from a patient focus group. Proc AMIA Annu Fall Symp 1997:672-676 [FREE Full text] [Medline: 9357710]
28. Duggan C, Bates I. Medicine information needs of patients: the relationships between information needs, diagnosis and disease. Qual Saf Health Care 2008 Apr;17(2):85-89. [doi: 10.1136/qshc.2005.017590] [Medline: 18385399]
29. Barney G. Doing Grounded Theory: Issues & Discussion. Mill Valley, CA: Sociology Press; 1998.
30. Coyne I. Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? J Adv Nurs 1997 Sep;26(3):623-630. [doi: 10.1046/j.1365-2648.1997.t01-25-00999.x] [Medline: 9378886]
31. Charmaz K. Teaching theory construction with initial grounded theory tools: a reflection on lessons and learning. Qual Health Res 2015 Dec;25(12):1610-1622. [doi: 10.1177/1049732315613982] [Medline: 26646825]
32. Smith R. What clinical information do doctors need? BMJ 1996 Oct 26;313(7064):1062-1068 [FREE Full text] [doi: 10.1136/bmj.313.7064.1062] [Medline: 8898602]
33. Anderson JG. Clearing the way for physicians' use of clinical information systems. Commun. ACM 1997 Aug;40(8):83-90. [doi: 10.1145/257874.257895]
34. Rumsey N, Clarke A, Musa M. Altered body image: the psychosocial needs of patients. Br J Community Nurs 2002 Nov;7(11):563-566. [doi: 10.12968/bjcn.2002.7.11.10886] [Medline: 12447117]
35. Julien H, Fourie I. Reflections of affect in studies of information behavior in HIV/AIDS contexts: An exploratory quantitative content analysis. Libr Inf 2015 Jan;37(1):3-9. [doi: 10.1016/j.lisr.2014.09.001]
36. O'Grady L. Meeting health information needs of people with HIV/AIDS: sources and means of collaboration. Health Info Libr J 2008 Dec;25(4):261-269 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00764.x] [Medline: 19076672]
37. Brashers DE, Goldsmith DJ, Hsieh E. Information seeking and avoiding in health contexts. Human Comm Res 2002 Apr;28(2):258-271. [doi: 10.1111/j.1468-2958.2002.tb00807.x]
38. Reay SD, Collier G, Douglas R, Hayes N, Nakarada-Kordic I, Nair A, et al. Prototyping collaborative relationships between design and healthcare experts: mapping the patient journey. Design for Health 2017 Mar 10;1(1):65-79. [doi: 10.1080/24735132.2017.1294845]
39. Trebble TM, Hansi N, Hydes T, Smith MA, Baker M. Process mapping the patient journey: an introduction. BMJ 2010 Aug 13;341:c4078. [doi: 10.1136/bmj.c4078] [Medline: 20709715]
40. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YY, Kenny LC, Adam F. Embedding the pillars of quality in health information technology solutions using "Integrated Patient Journey Mapping" (IPJM): case study. JMIR Hum Factors 2020 Sep 17;7(3):e17416 [FREE Full text] [doi: 10.2196/17416] [Medline: 32940610]
41. Kimbell L. The Service Innovation Handbook: Action-oriented Creative Thinking Toolkit for Service Organizations. Amsterdam: BIS Publishers; 2014.

42. Brown DL, Frank JE. Diagnosis and management of syphilis. Am Fam Physician 2003 Jul 15;68(2):283-290 [FREE Full text] [Medline: 12892348]
43. Ciria-Suarez L, Jiménez-Fonseca P, Palacín-Lois M, Antoñanzas-Basa M, Férnández-Montes A, Manzano-Fernández A, et al. Ascertaining breast cancer patient experiences through a journey map: A qualitative study protocol. PLoS One 2020;15(12):e0244355 [FREE Full text] [doi: 10.1371/journal.pone.0244355] [Medline: 33347475]
44. Nanton V, Docherty A, Meystre C, Dale J. Finding a pathway: information and uncertainty along the prostate cancer patient journey. Br J Health Psychol 2009 Sep;14(Pt 3):437-458. [doi: 10.1348/135910708X342890] [Medline: 18718111]
45. Mirnezami R, Kinross J, Vorkas P, Goldin R, Holmes E, Nicholson J, et al. Implementation of molecular phenotyping approaches in the personalized surgical patient journey. Ann Surg 2012 May;255(5):881-889. [doi: 10.1097/SLA.0b013e31823e3c43] [Medline: 22156927]
46. Naumer C, Fisher K. Information needs. In: Encyclopedia of Library and Information Sciences (3rd edition). Boca Raton, FL: CRC Press; 2010:2452-2458.
47. Fisher K, Erdelez S, McKechnie L. Theories of Information Behavior. Medford, NJ: Information Today, Inc; 2005.
48. Osorio MA. Youth information - seeking behavior II: context, theories, models and issues. Collection Building 2008 Apr 18;27(2):90. [doi: 10.1108/01604950810870254]
49. Tristram A. HPV information needs. Best Pract Res Clin Obstet Gynaecol 2006 Apr;20(2):267-277. [doi: 10.1016/j.bpobgyn.2005.10.006] [Medline: 16403488]
50. Okoniewski AE, Lee YJ, Rodriguez M, Schnall R, Low AFH. Health information seeking behaviors of ethnically diverse adolescents. J Immigr Minor Health 2014 Aug;16(4):652-660 [FREE Full text] [doi: 10.1007/s10903-013-9803-y] [Medline: 23512322]
51. Moskowitz D, Melton D, Owczarzak J. PowerON: the use of instant message counseling and the Internet to facilitate HIV/STD education and prevention. Patient Educ Couns 2009 Oct;77(1):20-26 [FREE Full text] [doi: 10.1016/j.pec.2009.01.002] [Medline: 19217742]

Abbreviations

ASHA: American Sexual Health Association
GT: grounded theory
LGBT: lesbian, gay, bisexual, and transgender
STI: sexually transmitted infection

Edited by R Kukafka, G Eysenbach; submitted 04.05.21; peer-reviewed by A Gupta, M Vică; comments to author 28.06.21; revised version received 27.07.21; accepted 04.10.21; published 10.11.21.

Please cite as:
Mulgund P, Sharman R, Purao S, Thimmanayakanapalya SS, Winkelstein P
Mapping Information Needs of Patients With Sexually Transmitted Infections Using Web-Based Data Sources: Grounded Theory Investigation
J Med Internet Res 2021;23(11):e30125
URL: https://www.jmir.org/2021/11/e30125
doi:10.2196/30125
PMID:34757326

Original Paper

Willingness of Chinese Men Who Have Sex With Men to Use Smartphone-Based Electronic Readers for HIV Self-testing: Web-Based Cross-sectional Study

Gifty Marley[1,2], MMed; Gengfeng Fu[3], PhD; Ye Zhang[4], MPH; Jianjun Li[3], MSc; Joseph D Tucker[2,5], MSc; Weiming Tang[2,6], PhD; Rongbin Yu[1], PhD

[1]School of Public Health, Nanjing Medical University, Nanjing, China

[2]The Social Entrepreneurship to Spur Health Project, The University of North Carolina Project-China, Guangzhou, China

[3]Section of STD/AIDS Prevention and Control, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[4]Kirby Institute, The University of New South Wales, Sydney, Australia

[5]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[6]Guangdong Second Provincial General Hospital, Guangzhou, China

Corresponding Author:
Rongbin Yu, PhD
School of Public Health
Nanjing Medical University
101 Longmian Avenue
Jiangning District
Nanjing, 211166
China
Phone: 86 13851545125
Email: rongbinyu@njmu.edu.cn

Abstract

Background: The need for strategies to encourage user-initiated reporting of results after HIV self-testing (HIVST) persists. Smartphone-based electronic readers (SERs) have been shown capable of reading diagnostics results accurately in point-of-care diagnostics and could bridge the current gaps between HIVST and linkage to care.

Objective: Our study aimed to assess the willingness of Chinese men who have sex with men (MSM) in the Jiangsu province to use an SER for HIVST through a web-based cross-sectional study.

Methods: From February to April 2020, we conducted a convenience web-based survey among Chinese MSM by using a pretested structured questionnaire. Survey items were adapted from previous HIVST feasibility studies and modified as required. Prior to answering reader-related questions, participants watched a video showcasing a prototype SER. Statistical analysis included descriptive analysis, chi-squared test, and multivariable logistic regression. *P* values less than .05 were deemed statistically significant.

Results: Of 692 participants, 369 (53.3%) were aged 26-40 years, 456 (65.9%) had ever self-tested for HIV, and 493 (71.2%) were willing to use an SER for HIVST. Approximately 98% (483/493) of the willing participants, 85.3% (459/538) of ever self-tested and never self-tested, and 40% (46/115) of unwilling participants reported that SERs would increase their HIVST frequency. Engaging in unprotected anal intercourse with regular partners compared to consistently using condoms (adjusted odds ratio [AOR] 3.04, 95% CI 1.19-7.74) increased the odds of willingness to use an SER for HIVST. Participants who had ever considered HIVST at home with a partner right before sex compared to those who had not (AOR 2.99, 95% CI 1.13-7.90) were also more willing to use an SER for HIVST. Playing receptive roles during anal intercourse compared to playing insertive roles (AOR 0.05, 95% CI 0.02-0.14) was associated with decreased odds of being willing to use an SER for HIVST. The majority of the participants (447/608, 73.5%) preferred to purchase readers from local Centers of Disease Control and Prevention offices and 51.2% (311/608) of the participants were willing to pay less than US $4.70 for a reader device.

Conclusions: The majority of the Chinese MSM, especially those with high sexual risk behaviors, were willing to use an SER for HIVST. Many MSM were also willing to self-test more frequently for HIV with an SER. Further research is needed to ascertain the diagnostic and real-time data-capturing capacity of prototype SERs during HIVST.

(J Med Internet Res 2021;23(11):e26480) doi:10.2196/26480

KEYWORDS

smartphone-based electronic reader; electronic readers; HIV self-testing; HIVST; self-testing; cellular phone–based readers; mHealth

Introduction

Many countries, including China, have adopted HIV self-testing (HIVST) strategies to complement HIV testing services following the World Health Organization's recommendation in 2016 [1]. Subsequently, various studies found the use of HIVST acceptable to key populations such as men who have sex with men (MSM) in China and around the world [2-5]. This strategy has also been shown to help reach hidden MSM who may have never accessed facility-based testing owing to fear of stigma and discrimination or other reasons [6,7]. Moreover, study participants have been cited to deem the privacy, convenience, and confidentiality associated with HIVST-facilitating factors for uptake, in addition to ease of use and short wait time [4,8].

HIVST continues to play a significant role in expanding HIV-testing services and was especially vital in retaining testing services during the COVID-19 outbreak when mobility was restricted [9-11]. For example, people in countries such as China, where HIVST has been approved for sale, could purchase kits online for doorstep delivery. This helped keep HIV testing services available and accessible during the pandemic period. However, upscaling HIVST to nationwide rollouts remains slow for many countries, including China, owing to existing gaps in distribution policies and integration strategies, persisting health worker concerns about low rates of user-initiated reporting of results, and lack of standardized reporting systems [12-15]. Although some studies found that the implementation of HIVST interventions where users can report their test results online (ie, texting the results to a health care worker on social media) was helpful to improve HIVST uptake and repeated HIV testing among Chinese MSM [16], reporting of posttest results remains largely dependent on provider-initiated routine follow-ups (such as home visits and reminder calling) [15,17-19]. Thus, the need for strategies that encourage user-initiated reporting of results at minimal/no extra cost to service providers requires further innovative measures.

Smartphone-based electronic readers (SERs) are smartphones adapted for health monitoring or clinical diagnostics and may include the use of external compatible hardware attachments (such as external optical lenses) to perform microscopic imaging [20,21]. SERs have become more popular in recent years because they consist of wearable heart rate monitors, temperature check apps, fitness products such as Fitbit watches, and blood glucose tracker apps [22,23]. Although the risks of private data leakage exist [24], the advantages SERs present for remote outpatient monitoring and early disease diagnosis cannot be understated [24,25]. For example, by using wearable sensors that feed into smartphone-based health apps, health care staff can monitor important physiological signs and activities of patients at home in real time from far-off facilities [26]. This could enable patients with chronic diseases to stay in care at home and is relatively cheaper compared to health facility admissions. Hence, SERs present an efficient cost-effective alternative to on-site clinical monitoring [27]. Furthermore, the emergence of SERs is vital to the expansion of decentralized infectious disease diagnostics [28], because unlike traditional laboratory diagnostics, they are portable and easy to use with minimal training required.

SERs are not a replacement for traditional laboratory diagnostics, but they could serve as preliminary diagnostic systems for early infectious disease diagnosis, especially in remote areas [20,21,29,30]. SERs such as smartphone-based bioimaging, smartphone-based bioanalyzers, and smartphone-based immune biosensors have been proven useful in point-of-care diagnostics [29,31,32]. Some pilot studies have also evaluated prototype SER use in point-of-care and found them to have comparable diagnostic capacity to traditional laboratory diagnostics. For example, a prototype smartphone-based rapid diagnostic test scanner was found capable of interpreting rapid diagnostic tests of 2 malaria species with comparable accuracy to standard visual interpretation [33,34]. Another study found a prototype SER capable of analyzing unprocessed liquid semen samples and providing users with a World Health Organization standard semen quality evaluation report with 98% accuracy [35]. Furthermore, other studies have shown prototype SERs capable of interpreting HIV rapid diagnostic test results with good accuracy and highlighted their capacity to capture real-time data [36,37]. Hence, coupled with recent advancements in smartphone computing, artificial intelligence, and open-source operating systems, SERs could be potential substitutes for bridging patient–health care provider barriers, and they could facilitate HIVST results reporting and active posttest counseling services through real-time data capturing.

However, as an emerging innovative tool without public implementation, it is not yet clear if HIVST users will be willing to use SERs for HIVST. Hence, there is a need to investigate public willingness to use SERs and to identify the determining factors associated with HIVST users' willingness to use SERs prior to adopting it. This study aimed to assess the willingness of MSM in the Jiangsu province of China to use an SER for HIVST through a web-based cross-sectional study.

Methods

Study Design

HIV prevalence among MSM in the Jiangsu province in the eastern part of China increased from 6.6% in 2015 [38] to about 8% in 2019, according to the Jiangsu Provincial Centers for Disease Control and Prevention (CDC) surveillance data. Between February and April 2020, we conducted a web-based cross-sectional study among Chinese MSM in the Jiangsu province. The survey was conducted using a structured questionnaire hosted on a web-based survey platform (wjx.cn).

Questions on sociodemographic data, sexual behaviors, and HIV testing experiences were adopted from a previous questionnaire used by our team in other MSM-related studies [39,40]. Questions on willingness to use smartphone-based readers were adopted from a previously used HIVST acceptability questionnaire and modified accordingly [41]. We pretested the questionnaire among 40 MSM purposively selected from a convenient sample of MSM clients who visited the Jiangsu Provincial CDC clinic for HIV testing services.

Convenience Web-Based Sampling

To participate in the study, participants who clicked the survey link had to satisfy some prespecified eligibility criteria, which included male sex at birth, be aged 16 years and older, had ever engaged in sex with other men, lived in the Jiangsu province, and be willing to provide informed consent to join the study. Willing participants who met the eligibility criteria had to agree to voluntarily participate in our study and for their provided data to be used strictly for research purposes. Participants who consented to this were redirected to the survey page to complete the questionnaire. The psychological state of the participants who met the eligibility criteria and provided informed consent was however not ascertained.

Data Collection

This study solicited information on participants' demographic variables, including their age, education level, marital status, household residency status, monthly income, and sexual orientation. Data solicited on sexual behaviors included sexual orientation disclosure to others, including sex workers, apart from sexual partners, number of sexual partners in the preceding 6 months, and usual places of meeting sexual partners (eg, pubs, night clubs, public bathrooms, online, through friends, others). Participants were also asked how often they used condoms during anal sex with their sexual partners in the preceding 6 months, if they ever participated in group sex/orgies (defined as sex with at least 2 men at the same time), and if they knew the HIV status of their sexual partners. We also collected information on the number of casual and regular male partner(s) in the preceding 6 months as well as the roles participants played during anal sex with regular and casual partners. HIV testing history was assessed with the questions: how often do you test for HIV, how many times did you test for HIV in the past year, how many times have you tested for HIV this year, do you know your HIV status, do you know the HIV status of your regular partner, and have you ever disclosed your HIV status to any of your sexual partners. Participants were also asked if they had ever self-tested for HIV, where they obtained HIVST kits, if they used HIVST for their first HIV testing, and if they had ever considered testing for HIV at home with a partner right before sex. Information on self-reported concerns about HIVST as well as willingness to report HIVST results by sending pictures of the used test kits to health care workers or researchers were also collected.

Assessment of Willingness

Participants watched a short video showcasing a prototype SER prior to answering questions on willingness (Multimedia Appendix 1), in addition to a short introductory paragraph on SERs to provide context. The prototype consisted of a portable 3D printed case, which housed a light-emitting diode light source, a mini scanner, a Bluetooth tag, and a closing cap to ensure stability and safe hold during use. The detailed description of the prototype SER will be made available upon completion.

SER Use

Self-testers need to connect the reader to a smartphone device via Bluetooth after turning on the reader. After installing the needed software and entering some basic biodata, users can proceed to uncap the reader, place the used HIVST kit into the reader slot following the direction (to know which part to insert), and close the cap to ensure the kit is secured in place. The reader would then scan the result display area of the test kit by using the embedded light-emitting diode scanner and highlight the display of the test lines (red strips) by reducing the background color value. This would enable the reader software to check whether the control line (C) and the test line (T) are displayed or not. The software will interpret the results as negative (C + T -), positive (C + T +), or invalid (C-T + / -) depending on which test lines are displayed. The reader software then transmits the results to a centralized system, which in turn feeds back the results to the smartphone and the CDC surveillance system concurrently. Participants were asked if they would be willing to use an SER for HIVST if it was provided for free. Subsequent questions asked if participants would recommend SERs to a sexual partner and whether having an SER would influence their HIVST frequency. Additional questions assessed participant opinions on factors that would facilitate and discourage them from opting to use SERs in future as well as how much they would be willing to pay to purchase such a reader device.

Statistical Analysis

A descriptive analysis was conducted to summarize the sociodemographic characteristics, sexual behaviors, and HIV testing experience. Chi-squared test was used to test for statistically significant factors associated with MSM willingness to use smartphone-based HIVST readers if offered during self-testing. Individual correlation models were used to assess variables' association with willingness to use an SER for HIVST. Factors that had a significant correlation with the dependent variable were entered into a multivariable logistic regression. Age, income, marital status, and level of education were considered potential confounders and adjusted in the model [8,42-44]. Statistical significance was defined at P values <.05, and all analyses were conducted using SPSS statistics version 23.0.0 (IBM Corp).

Ethical Approval

Ethical approval was obtained from the ethics review committees at the Jiangsu Provincial CDC (project JSJK2019-B016-03). All participants provided virtual informed consent prior to participation in the study.

Results

Sociodemographic Characteristics and HIV Testing History

A total of 692 completed surveys were included in the data analysis. The majority of the participants (369/692, 53.3%) were aged between 26 and 40 years (median age 31-40 years), 30.5% (211/692) were older than 40 years, and 16.2% (112/692) were 25 years old and younger. Of all the participants, 67.5% (467/692) self-identified as homosexuals, 51.3% (355/692) had a college or university degree, and 50.9% (352/692) earned more than US $780 monthly. Most participants had disclosed their sexual orientation to others apart from their sexual partners (552/692, 79.8%), and the majority met sexual partners through web-based platforms (594/692, 85.8%). Many participants (437/692, 63.2%) had at least one regular sexual partner and 26% (180/692) reported having multiple temporary sexual partners in the preceding 6 months. Unprotected anal intercourse was more frequent with regular than with causal partners in the last 6 months (184/692, 26.6% vs 75/692,10.8%, respectively) (Table 1).

Table 1. Descriptive analysis of the sociodemographic characteristics and sexual history of men who have sex with men in China in 2020 (N=692).

Variables	Values, n (%)
Age (years)	
≤25	112 (16.2)
26-40	369 (53.3)
>40	211 (30.5)
Marital status	
Single	340 (49.1)
Married	239 (34.5)
Divorced/separated	113 (16.3)
Highest education level	
High school	295 (42.6)
College/University	355 (51.3)
Postgraduate	42 (6.1)
Monthly income (USD)	
<230	28 (4)
230-780	312 (45.1)
>780	352 (50.9)
Sexual identity	
Homosexual	467 (67.5)
Bisexual	211 (30.5)
Unsure	14 (2)
Disclosed sexual orientation to others apart from sexual partners	
Yes	552 (79.8)
No	140 (20.2)
Places you usually meet sexual partners	
Hotspots (ie, pubs, bath houses, massage parlor, etc)	142 (20.5)
Public places (eg, parks, public restroom)	56 (8.1)
Online (ie, dating websites, social media)	594 (85.8)
Met through family/friends	180 (26)
Don’t search sexual partners	87 (12.6)
Unprotected anal intercourse in the last 6 months	
Yes	202 (29.2)
No	392 (56.6)
Have had no sexual partners	98 (14.2)
Number of regular sexual partners in the last 6 months	
None	255 (36.8)
1	348 (50.3)
≥2	89 (12.9)
Role during sex with regular partner	
Insertive (n=437)	232 (53.1)
Both (n=437)	131 (30)
Receptive (n=437)	74 (16.9)
Had no regular sexual partners (N=692)	255 (36.8)

Variables	Values, n (%)
Unprotected anal intercourse with regular partners	
Yes (n=437)	184 (42.1)
No (n=437)	253 (57.9)
Had no regular sexual partners (N=692)	255 (36.8)
Number of temporary partners in the last 6 months	
None	378 (54.6)
1	134 (19.4)
≥2	180 (26)
Role during sex with temporary partner	
Insertive (n=314)	111 (35.4)
Both (n=314)	69 (22)
Receptive (n=314)	84 (26.8)
Had no casual partners (N=692)	378 (54.6)
Refused to answer (n=314)	50 (15.9)
Unprotected anal intercourse with casual partner	
Yes (n=314)	75 (28.4)
No (n=314)	189 (71.6)
Had no casual partners (N=692)	378 (54.6)
Refused to answer (n=314)	50 (15.9)

Most participants (495/692, 71.5%) had tested for HIV more than twice in the past year; 65.9% (456/692) had ever used an HIV self-test kit, of which 42.5% (194/456) used it for their first HIV test. The majority of the participants (560/692, 80.9%) had also considered home testing with their partner right before sex. Table 2 shows further descriptive details of participants' HIV testing history.

Table 2. Descriptive analysis of participants' HIV testing habits and self-testing experience in China in 2020.

Variables	Values, n (%)
How often do you test for HIV? (N=692)	
3 months	398 (57.5)
6 months	140 (20.2)
1 year or more	154 (22.3)
Number of times tested for HIV last year (N=692)	
Never	80 (11.6)
Once	117 (16.9)
≥Twice	495 (71.5)
Know your HIV status (N=692)	
Yes	642 (92.8)
No	50 (7.2)
Ever self-tested for HIV (N=692)	
Yes	456 (65.9)
No	156 (22.5)
Never heard of self-testing	80 (11.6)
Used HIV self-testing kit for first HIV test	
Yes (n=456)	194 (42.5)
No (n=456)	262 (57.5)
Ever considered testing for HIV at home prior to a sexual encounter (N=692)	
Yes	560 (80.9)
No	132 (19.1)
Concerns when you use HIV self-testing (N=692)	
Is it working correctly?	237 (34.2)
Time to wait for results to appear	204 (29.5)
Results accuracy	201 (29)
Read the results correctly	276 (39.9)
Comfortable sharing a picture of HIVST results with health worker for reporting (N=692)	
Yes	347 (50.1)
No	345 (49.9)

Factors Associated With Willingness to Use an SER

Of the 692 participants, 493 (71.2%) were willing to use an SER during HIVST, 115 (16.6%) were unwilling, and 84 (12.1%) were unsure of their willingness. The majority of the willing participants (428/493, 86.8%) and few of unwilling participants (45/115, 39.1%) agreed they would recommend SERs to their sexual partners for HIVST. Additionally, 98% (483/493) of the participants agreed that having an SER would increase their HIVST frequency. Willing participants cited the following factors as the major facilitators that would encourage their use of an SER: obtaining accurate self-test results (328/493, 66.5%), ease of use (300/493, 60.9%), and short wait time of 15-20 minutes for results (251/493, 50.9%). Alternatively, cost of the reader (259/493, 52.5%) and fear of test results leaking to others (214/493, 43.4%) were deemed barriers. Most willing participants preferred to purchase SERs from local offices of the CDC (375/493, 76.1%), pharmacies/supermarkets (200/493, 40.6%), and vending machines (192/493, 38.9%). The acceptable cost for an SER was less than US $4.30 (239/493, 48.5%), although 32% (158/493) were willing to pay US $8.70 or more. Among the 18.9% (115/692) unwilling participants, 61.7% (71/115) would consider using SERs owing to its ease of use and less wait time of 15-20 minutes for results (57/115, 49.6%), while having never heard of the reader (59/115, 51.3%) and purchase cost (43/115, 37.4%) would deter use. Likewise, many unwilling participants preferred to buy SERs from local CDC offices (72/115, 62.6%) and hospitals/clinics (60/115, 52.2%) at an acceptable cost of less than US $4.30 (72/115, 62.6%) (Table 3).

Table 3. Factors associated with willingness to use smartphone-based electronic readers among Chinese men who have sex with men.

Variables	Willing to use reader		Total (n=608), n (%)	*P* value
	Yes (n=493), n (%)	No (n=115), n (%)		
Would recommend HIV self-testing kit reader to sexual partner				<.001[a]
Yes	428 (86.8)	45 (39.1)	473 (77.8)	
No	65 (13.2)	70 (60.9)	135 (22.2)	
HIV self-testing kit reader would increase your HIV testing frequency				<.001[a]
Yes	483 (98)	46 (40)	529 (87)	
No	10 (2)	69 (60)	79 (13)	
Concerned about reading HIV self-testing results correctly				.001[b]
Yes	174 (35.3)	60 (52.2)	234 (38.5)	
No	319 (64.7)	55 (47.8)	374 (61.5)	
Facilitators of smartphone-based electronic reader use				
Easy to use	300 (60.9)	71 (61.7)	371 (61)	.86
You don't need to interpret the results of HIV self-testing yourself	213 (43.2)	11 (9.6)	224 (40.1)	<.001[a]
15-20 minutes to find out the results of HIV self-testing	251 (50.9)	57 (49.6)	289 (47.5)	.79
Get accurate results	328 (66.5)	46 (40)	374 (61.5)	<.001[a]
Referral services can be implemented online if the test results are positive	200 (40.6)	49 (42.6)	249 (41)	.69
Barriers to smartphone-based electronic reader use				
Never heard of it	176 (35.7)	59 (51.3)	235 (38.7)	.002[b]
Worried about accuracy	141 (28.6)	25 (21.7)	166 (27.3)	.14
Cost	259 (52.5)	43 (37.4)	302 (49.7)	.003[b]
Worried about test results leak to others (including to health care workers)	214 (43.4)	42 (36.5)	256 (42.1)	<.001[a]
Worried that others will see the test results on the screen or on their phone	178 (36.1)	31 (27)	209 (34.4)	.002[b]
Worried about personal data being misused	187 (37.9)	26 (22.6)	213 (35)	<.001[a]
Waiting anxiety and fear of positive results	105 (21.3)	N/A[c]	105 (17.3)	
Where will you like to obtain a self-test reader				
Pharmacy, supermarket, etc	200 (40.6)	11 (9.6)	211 (34.7)	<.001[a]
Web-based mall	167 (33.9)	54 (47)	221 (36.3)	.009[b]
Hospital or clinic	153 (31)	60 (52.2)	213 (35)	<.001[a]
Vending machines	192 (38.9)	45 (39.1)	237 (39)	.97
Local Centers for Disease Control and Prevention office	375 (76.1)	72 (62.6)	447 (73.5)	.003
How much are you willing to pay for a HIV self-testing kit reader (USD)				.003[b]
<4.60	239 (48.5)	72 (62.6)	311 (51.2)	
4.70-9.40	96 (19.5)	16 (13.9)	112 (18.4)	
>9.50	158 (32)	27 (23.5)	185 (30.4)	

[a]Statistically significant at *P*<.01

[b]Statistically significant at *P*<.05.

[c]N/A: not applicable.

Furthermore, 78.3% (299/382) of the participants who had "ever self-tested" for HIV and 79.3% (124/156) "never self-tested" participants were willing to use SERs. The majority of the "ever self-tested" and "never self-tested" participants (311/382, 81.4% and 124/156, 79.5%, respectively) would recommend the SERs to their partners and agreed that having an SER would increase their HIVST frequency (335/382, 87.7% and 124/156, 79.5%, respectively). Willing "ever self-tested participants" deemed SERs' wait time of 15-20 minutes (263/382, 68.8%), ease of use (249/382, 65.2%), and result accuracy (233/382, 61%) to be facilitating factors. Cost of reader (214/382, 56%), worry about test results leaking to others (204/382, 53.4%), and concerns about personnel data being misused (148/382, 38.7%) were deemed barriers.

Among the "never self-tested" participants, however, obtaining accurate results (92/156, 59%) and ease of use (84/156, 53.8%) were key facilitators, while having never heard of SERs (83/156, 53.2%), worry about test results leaking to others (54/156, 34.6%), and concerns about personnel data being misused (54/156, 34.6%) were deemed barriers. The majority of both "self-tested" and "never self-tested" participants preferred to buy SERs from their local CDC offices (293/382, 76.7% and 124/156, 79.5%, respectively), and 48.3% (260/538) were willing to pay less than 30 RMB (approximately US $4.60) for a reader (Table 4).

Table 4. Factors associated with the willingness of ever self-tested and never self-tested Chinese men who have sex with men to use smartphone-based electronic readers for HIV self-testing.

Variables	Self-tested (n=382), n (%)	Never self-tested (n =156), n (%)	Total (n=538), n (%)	*P* value
Would recommend HIV self-testing kit reader to sexual partner				.63
Yes	311 (81.4)	124 (79.5)	435 (80.9)	
No	60 (15.7)	32 (20.5)	92 (17.1)	
Unsure	11 (2.9)	N/A[a]	11 (2)	
HIV self-testing kit reader would increase your HIV testing frequency				.02[b]
Yes	335 (87.7)	124 (79.5)	459 (85.3)	
No	11 (2.9)	32 (20.5)	43 (8)	
Unsure	36 (9.4)	N/A	36 (6.7)	
Facilitators of HIV self-testing reader use				
Easy to use	249 (65.2)	84 (53.8)	333 (61.9)	.01[b]
You don't need to interpret the results of HIV self-testing yourself	155 (40.6)	52 (33.3)	207 (38.5)	.12
15-20 minutes to find out the results of HIV self-testing	263 (68.8)	28 (17.9)	291 (54.1)	.001[c]
Get accurate results	233 (61)	92 (59)	325 (60.4)	.66
Referral services can be implemented online if the test is reactive	173 (45.3)	46 (29.5)	219 (40.7)	.001[b]
Barriers to HIV self-testing reader use				<.001[c]
Never heard of it	135 (35.3)	83 (53.2)	218 (40.5)	
Worried about accuracy	127 (33.2)	26 (16.7)	153 (28.4)	
Cost of device	214 (56)	54 (34.6)	268 (49.8)	
Worry about test results leaking to others (including to health workers)	204 (53.4)	54 (34.6)	258 (48)	
Worry about personal data being misused	148 (38.7)	33 (21.2)	181 (33.6)	
Waiting anxiety and fear of positive results	105 (27.5)	N/A	105 (19.5)	
Where will you like to obtain a self-test reader?				
Shops (pharmacy, supermarket, mall, etc)	99 (25.9)	61 (39.1)	160 (29.7)	.04[b]
Web-based mall	139 (36.4)	33 (21.2)	172 (32)	.001[b]
Hospital or clinic	117 (30.6)	79 (50.6)	196 (36.4)	<.001[c]
Vending machines	124 (32.5)	77 (49.4)	201 (37.4)	<.001[c]
Local Centers for Disease Control and Prevention office	293 (76.7)	124 (79.5)	417 (77.5)	.57
How much are you willing to pay for a HIV self-testing kit reader? (USD)				<.001[c]
<4.60	174 (45.5)	86 (55.1)	260 (48.3)	
4.70-9.40	58 (15.2)	54 (34.6)	112 (20.8)	
≥9.50	150 (39.3)	16 (10.3)	166 (30.9)	

[a]N/A: not applicable.

[b]Correlation significant at $P<.05$.

[c]Correlation significant at $P<.01$.

Predictors of MSM Willingness to Use an SER

MSM participants who engaged in unprotected anal intercourse with regular partners compared to those who consistently used condoms (adjusted odds ratio [AOR] 3.04, 95% CI 1.19-7.74) were more likely to be willing to use SERs for HIVST. Participants who had ever considered self-testing for HIV at home with a partner right before sex compared to those who had not (AOR 2.99, 95% CI 1.13-7.90) were also more willing to use an SER for HIVST. However, participants who played receptive roles during anal intercourse compared to those who played insertive roles (AOR 0.05, 95% CI 0.02-0.14) were less likely to be willing to use an SER for HIVST. Although not statistically significant, participants older than 40 years compared to participants 25 years old and younger (AOR 1.57, 95% CI 0.50-4.94), participants with a postgraduate level of education compared to participants who completed only high school (AOR 1.22, 95% CI 0.31-4.73), and participants who tested for HIV every 3 months compared to those who tested once a while (AOR 2.44, 95% CI 0.95-6.31) were also more willing to use an SER for HIVST. Table 5 further summarizes the details on predicting the factors associated with the willingness of MSM to use SERs.

Table 5. Logistic regression analysis of the factors predicting willingness to use smartphone-based electronic readers.

Variables	Total (n=608), n (%)	Willing to use readers		Adjusted odds ratio[a] (95% CI)	*P* value
		Yes (n=493), n (%)	No (n=115), n (%)		
Demographics characteristics					
Age (years)					
26-40	330 (54.3)	258 (52.3)	72 (62.6)	0.61 (0.24-1.52)	.29
>40	187 (30.8)	163 (33.1)	24 (20.9)	1.57 (0.50-4.94)	.44
≤25	91 (15)	72 (14.6)	19 (16.5)	1	
Marital status					
Single	296 (48.7)	234 (47.5)	62 (53.9)	1.81 (0.68-4.76)	.23
Married	215 (35.4)	181 (36.7)	34 (29.6)	2.13 (0.81-5.63)	.13
Divorced/separated	97 (16)	78 (15.8)	19 (16.5)	1	
Highest education level					
College/University	314 (51.6)	256 (51.9)	58 (50.4)	0.69 (0.32-1.45)	.32
Postgraduate	40 (6.6)	32 (6.5)	8 (7)	1.22 (0.31-4.73)	.78
High school	254 (41.8)	205 (41.6)	49 (42.6)	1	
Monthly income (USD)					
<230	24 (3.9)	17 (3.4)	7 (6.1)	2.23 (0.32-15.48)	.42
230-780	273 (44.9)	225 (45.6)	48 (41.7)	1.06 (0.53-2.10)	.87
>780	311 (51.2)	251 (50.1)	60 (52.2)	1	
Sexual behaviors					
Role during sex with regular partner in the last 6 months					
Both	168 (27.6)	120 (24.3)	48 (41.7)	0.99 (0.39-2.51)	.98
Receptive	101 (16.6)	41 (8.3)	60 (52.2)	0.05 (0.02-0.14)	<.001[b]
Insertive	185 (30.4)	169 (34.3)	16 (13.9)	1	
Unprotected anal intercourse with regular partner in the last 6 months					
Yes	173 (28.5)	155 (31.4)	18 (15.7)	3.04 (1.19-7.74)	.02[c]
No	364 (59.9)	278 (56.4)	86 (74.8)	1	
Ever considered self-testing at home prior to sex					
Yes	476 (78.3)	402 (81.5)	74 (64.3)	2.99 (1.13-7.90)	.03[c]
No	132 (21.7)	91 (18.5)	41 (35.7)	1	
HIV testing history					
How often do you test for HIV?					
3 months	338 (55.6)	285 (57.8)	53 (46.1)	2.44 (0.95-6.31)	.07
6 months	130 (21.4)	98 (19.9)	32 (27.8)	0.47 (0.17-1.33)	.16
1 year or more	140 (23)	110 (22.3)	30 (26.1)	1	

[a]Adjusted for age, income, marital status, and level of education.

[b]Significant at *P*<.01.

[c]Statistically significant at *P*<.05.

Discussion

Principal Results

User-initiated reporting of results and linkage to care after self-testing are persistent barriers to HIVST uptake and scale-up. This study extends existing literature on the use of SERs as it explores MSM opinions about SERs and their willingness to use it for HIVST. Our findings showed that the majority of the Chinese MSM were willing to use SERs for HIVST. In addition, MSM with high sexual risk behaviors and those who had ever considered self-testing were more willing to use an SER for HIVST. Furthermore, SERs could help increase HIVST uptake and increase testing frequency among both ever self-tested and never self-tested MSM. Finally, most MSM were willing to pay less than US $4.30 for an SER and preferred to obtain it from their local CDC office.

We found that most participants were willing to use an SER for HIVST. Our findings showed that the majority of MSM (493/608, 81.1%), including 78.3% (299/382) of ever self-tested MSM and 79.5% (124/156) of never self-tested participants, were willing to use an SER for HIVST. Our finding concurs with findings of a recent study that found the use of a prototype SER for HIVST acceptable to MSM and transgender women participants [45]. The study attributed the high acceptability to the capacity of the prototype SER to save as well as share HIVST results with partners and health care providers [45]. We, however, speculate that the observed high willingness in our study may be due to HIVST users' need to correctly read test results with assured accuracy. This explanation is plausible as 39.9% (276/692) of the participants in our study reported having concerns about their ability to accurately read and understand self-test results. Although this is a preliminary study, our findings provide some foundational evidence to trigger further research on the role of SERs in bridging HIVST-related gaps and expanding HIVST uptake. There is also a need for future studies to evaluate the real-time data-capturing capacity of prototype SERs and assess their functional integration into existing health systems.

In addition, MSM with higher sexual risk behaviors were more willing to use an SER for HIVST. Compared to participants who consistently used condoms, MSM who engaged in unprotected anal intercourse were more willing to use an SER for HIVST. This could be because MSM with higher sexual risk behaviors test for HIV more frequently owing to a perceived sense of being at higher risk for HIV infection [17,46]. Further, in support of this explanation, previous studies have found MSM who report inconsistent condom use with sexual partners to be more likely to have tested for HIV recently and opt for self-testing [4,47]. We also speculate that frequent testers are more likely to opt for HIVST for convenience and privacy. Therefore, having an SER to accurately interpret results will be deemed an added advantage. We also found that MSM who had ever considered self-testing at home with a partner right before sexual intercourse were willing to use an SER for HIVST. Therefore, it is possible that having a perceived sense of high risk to HIV and wanting to use HIVST plays a role in determining MSM willingness to use SERs. Further research is needed to better understand the dynamics of factors that predict willingness to use SERs among MSM and other HIVST users. We also recommend further research into behavioral factors that predict willingness to use SERs among different populations to inform the promotion of SERs in future.

Our findings showed that SERs could improve HIVST uptake among MSM. The majority of the "ever self-tested" and "never self-tested" participants in our study reported that having a reader would increase their HIVST frequency. However, less than half of the "unwilling" participants (46/115, 40%) reported the same. Nonetheless, this is still an important observation as our findings showed a current HIVST uptake rate of 65.8% (456/692) among the participants. This rate is still below the optimal coverage although it is an increase from the observed 37.2% in 2014 [48]. In addition, although many studies have proven users capable of properly undertaking HIVST [4,6,8], user concerns about their capacity to accurately read and interpret results still persist [15,49]. Therefore, this technology presents an opportunity to expand HIV testing among MSM and should be further investigated. For the majority of the unwilling participants (59/115, 51.3%) and never self-tested participants (83/156, 53.2%), never having heard of SERs could be a major barrier to use. Therefore, we recommend that public education on SER be undertaken and information about SERs be made readily available and accessible prior to SER introduction. Policies that standardize emerging HIVST reporting systems to ensure their smooth integration into existing health reporting structures and guard against personal data abuse are also needed.

The cost and place of purchase are vital to the promotion of SERs. Findings from our study showed that 51.2% (311/608) of MSM were unwilling to pay more than US $4.60 for an SER. Similarly, the majority of both never self-tested participants (86/156, 55.1%) and unwilling participants (72/115, 62.6%) were also unwilling to pay more than US $4.30 for an SER. This result is similar to the findings of previous studies that have shown purchase cost to be a deterring barrier to HIVST uptake [6,50]. The majority of the participants also preferred purchasing SERs from local CDC offices. This concurs with observations from previous HIVST acceptability studies in China [41,51]. We could attribute this to the notion that Chinese MSM trust the CDC owing to their involvement in the delivery of key population–centered HIV interventions [52,53]. It is also possible that the recruiting community-based organizations' existing partnership with the CDC may have contributed to this preference. Therefore, in addition to local CDC offices, SERs should be made available at clinics, pharmacies, and web-based shops to facilitate accessibility to hidden populations when they are adopted. Furthermore, government agencies could team up with device manufacturers to minimize manufacturing cost and subsidize purchase cost to encourage uptake. Further, smartphone manufacturers should consider incorporating rapid tests reading functions and apps into emerging phone gadgets to minimize, if not eliminate, the extra cost of obtaining external SERs.

Our study has many implications as it evaluates the willingness of self-testers to use an emerging tool of importance to HIVST. Our findings highlight lay users' expectations of SERs in HIVST

that could inform manufacturers of specifications that will enhance SERs for better integration into HIVST programs. For research, future randomized controlled studies should assess the feasibility and acceptability of SERs among other population subtypes by using larger sample sizes. Furthermore, studies that seek to evaluate the sensitivity and specificity of SERs during HIVST by users should be conducted. Our study findings also support recommendations of previous studies that SERs should be embedded with cloud-based data functions. We also showcase the need for further evaluation of SER uses in real-time data capturing and monitoring. Lastly, quality assurance policies and guidelines are needed to inform the incorporation of SERs into HIVST programming as well as to standardize the manufacture and distribution of SERs.

Limitations

Our study has some limitations. First, as this was a web-based cross-sectional study, the MSM sample size may be unrepresentative of the larger Jiangsu MSM population. Further, owing to the convenience sampling method employed, our study findings may be ungeneralizable. However, our findings still serve as preliminary data to further guide future research on the acceptability of SERs for HIVST. Second, as participants had limited familiarity with SERs and the query assessed willingness on the preconditioned clause that SER will be provided for free, the results on the willingness to use it and secondary distribution may have been biased. Nonetheless, our findings showed that MSM were willing to purchase SERs and highlighted important features required to encourage SER use in HIVST programming. Third, our investigation addressed sensitive questions that may have led to the misreporting of personal sexual risk and HIV testing behavior data owing to social desirability bias. Lastly, the mental status and capacity of eligible participants who provided informed consent were not assessed prior to administering the survey. Therefore, we recommend further studies to be conducted to explore the acceptance of the HIVST reader results among other subpopulation types prior to public implementation.

Conclusion

Regardless of the limited knowledge, many Chinese MSM, especially those with high sexual risk behaviors, are willing to use SERs for HIVST. In addition, the majority of both ever self-tested and never self-tested MSM are willing to self-test more frequently with SERs. Therefore, SERs could facilitate HIVST uptake and scale-up among MSM. However, appropriate pricing and safe and anonymous procurement venues are key to facilitating SER uptake among key populations. Further research is needed to validate the uptake of SERs for HIVST among other key population subtypes.

Acknowledgments

We received support from the National Key Research and Development Program of China (2017YFE0103800), the National Institutes of Health (R34MH119963), National Science and Technology Major Project (2018ZX10101-001-001-003), National Natural Science Foundation of China (81903371), and the National Social Science Fund of China (19CSH018).

Authors' Contributions

WT, GF, and RY conceptualized the study. WT, GM, YZ, and JL designed the study and collected the data. GM, LJ, and YZ collated and statistically analyzed the data. GM and GF reviewed the literature and wrote the initial draft. WT, YZ, LJ, JDT, and RY provided critical revisions and references. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prototype smartphone-based electronic reader in use.
[MP4 File (MP4 Video), 9652 KB - jmir_v23i11e26480_app1.mp4]

References

1. Guidelines on HIV self-testing and partner notification: supplement to consolidated guidelines on HIV testing services. World Health Organization. URL: https://www.who.int/hiv/pub/vct/Annex19.pdf [accessed 2020-09-25]
2. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al. Uptake, Accuracy, Safety, and Linkage into Care over Two Years of Promoting Annual Self-Testing for HIV in Blantyre, Malawi: A Community-Based Prospective Study. PLoS Med 2015 Sep;12(9):e1001873 [FREE Full text] [doi: 10.1371/journal.pmed.1001873] [Medline: 26348035]
3. Hatzold K, Gudukeya S, Mutseta MN, Chilongosi R, Nalubamba M, Nkhoma C, et al. HIV self-testing: breaking the barriers to uptake of testing among men and adolescents in sub-Saharan Africa, experiences from STAR demonstration projects in Malawi, Zambia and Zimbabwe. J Int AIDS Soc 2019 Mar;22 Suppl 1:e25244 [FREE Full text] [doi: 10.1002/jia2.25244] [Medline: 30907505]
4. Wang XF, Wu ZY, Tang ZZ, Nong QX, Li YQ. [Acceptability of HIV testing using oral quick self-testing kit in men who have sex with men]. Zhonghua Liu Xing Bing Xue Za Zhi 2018 Jul 10;39(7):937-942. [doi: 10.3760/cma.j.issn.0254-6450.2018.07.014] [Medline: 30060308]

5. Tun W, Vu L, Dirisu O, Sekoni A, Shoyemi E, Njab J, et al. Uptake of HIV self-testing and linkage to treatment among men who have sex with men (MSM) in Nigeria: A pilot programme using key opinion leaders to reach MSM. J Int AIDS Soc 2018 Jul;21 Suppl 5:e25124 [FREE Full text] [doi: 10.1002/jia2.25124] [Medline: 30033680]
6. Liu F, Qin Y, Meng S, Zhang W, Tang W, Han L, et al. HIV self-testing among men who have sex with men in China: a qualitative implementation research study. J Virus Erad 2019 Nov 04;5(4):220-224 [FREE Full text] [Medline: 31754445]
7. Lyons CE, Coly K, Bowring AL, Liestman B, Diouf D, Wong VJ, et al. Use and Acceptability of HIV Self-Testing Among First-Time Testers at Risk for HIV in Senegal. AIDS Behav 2019 Sep;23(Suppl 2):130-141 [FREE Full text] [doi: 10.1007/s10461-019-02552-2] [Medline: 31197701]
8. Liu Y, Wu G, Lu R, Ou R, Hu L, Yin Y, et al. Facilitators and Barriers Associated with Uptake of HIV Self-Testing among Men Who Have Sex with Men in Chongqing, China: A Cross-Sectional Survey. Int J Environ Res Public Health 2020 Mar 03;17(5):1634 [FREE Full text] [doi: 10.3390/ijerph17051634] [Medline: 32138263]
9. Considerations for HIV self-testing in the context of the COVID-19 pandemic and its response: an operational update. Unitaid-STAR Initiative-ATLAS Project. URL: https://www.psi.org/wp-content/uploads/2020/06/Considerations-for-HIV-Self-Testing-in-the-Context-of-the-COVID-19-Pandemic-and-Its-Response_FIN.pdf [accessed 2020-08-15]
10. HIV testing strategy adaptations due to the COVID-19 pandemic attract new clients. USAID. URL: https://www.usaid.gov/vietnam/program-updates/may-2020-hiv-testing-strategy-adaptations-due-covid-19-pandemic-attract-new-clients [accessed 2020-08-12]
11. Hoagland B, Torres TS, Bezerra DRB, Benedetti M, Pimenta C, Veloso VG, et al. High acceptability of PrEP teleconsultation and HIV self-testing among PrEP users during the COVID-19 pandemic in Brazil. Braz J Infect Dis 2021;25(1):101037 [FREE Full text] [doi: 10.1016/j.bjid.2020.11.002] [Medline: 33285137]
12. Dzinamarira T, Kamanzi C, Mashamba-Thompson TP. Key Stakeholders' Perspectives on Implementation and Scale up of HIV Self-Testing in Rwanda. Diagnostics (Basel) 2020 Apr 01;10(4):194 [FREE Full text] [doi: 10.3390/diagnostics10040194] [Medline: 32244566]
13. Indravudh PP, Choko AT, Corbett EL. Scaling up HIV self-testing in sub-Saharan Africa: a review of technology, policy and evidence. Curr Opin Infect Dis 2018 Feb;31(1):14-24 [FREE Full text] [doi: 10.1097/QCO.0000000000000426] [Medline: 29232277]
14. Tahlil KM, Ong JJ, Rosenberg NE, Tang W, Conserve DF, Nkengasong S, et al. Verification of HIV Self-Testing Use and Results: A Global Systematic Review. AIDS Patient Care STDS 2020 Apr;34(4):147-156 [FREE Full text] [doi: 10.1089/apc.2019.0283] [Medline: 32324482]
15. Sharma M, Ying R, Tarr G, Barnabas R. Systematic review and meta-analysis of community and facility-based HIV testing to address linkage to care gaps in sub-Saharan Africa. Nature 2015 Dec 03;528(7580):S77-S85 [FREE Full text] [doi: 10.1038/nature16044] [Medline: 26633769]
16. Chan PS, Chidgey A, Lau J, Ip M, Lau JTF, Wang Z. Effectiveness of a Novel HIV Self-Testing Service with Online Real-Time Counseling Support (HIVST-Online) in Increasing HIV Testing Rate and Repeated HIV Testing among Men Who Have Sex with Men in Hong Kong: Results of a Pilot Implementation Project. Int J Environ Res Public Health 2021 Jan 15;18(2):729 [FREE Full text] [doi: 10.3390/ijerph18020729] [Medline: 33467770]
17. Jin X, Xu J, Smith MK, Xiao D, Rapheal ER, Xiu X, et al. An Internet-Based Self-Testing Model (Easy Test): Cross-Sectional Survey Targeting Men Who Have Sex With Men Who Never Tested for HIV in 14 Provinces of China. J Med Internet Res 2019 May 15;21(5):e11854 [FREE Full text] [doi: 10.2196/11854] [Medline: 31094339]
18. Lightfoot MA, Campbell CK, Moss N, Treves-Kagan S, Agnew E, Kang Dufour M, et al. Using a Social Network Strategy to Distribute HIV Self-Test Kits to African American and Latino MSM. J Acquir Immune Defic Syndr 2018 Sep 01;79(1):38-45. [doi: 10.1097/QAI.0000000000001726] [Medline: 29771792]
19. Mulubwa C, Hensen B, Phiri MM, Shanaube K, Schaap AJ, Floyd S, HPTN 071 (PopART) Study Team. Community based distribution of oral HIV self-testing kits in Zambia: a cluster-randomised trial nested in four HPTN 071 (PopART) intervention communities. Lancet HIV 2019 Feb;6(2):e81-e92 [FREE Full text] [doi: 10.1016/S2352-3018(18)30258-3] [Medline: 30584047]
20. Roda A, Michelini E, Zangheri M, Di Fusco M, Calabria D, Simoni P. Smartphone-based biosensors: A critical review and perspectives. TrAC Trends in Analytical Chemistry 2016 May;79(4):317-325 [FREE Full text] [doi: 10.1016/j.trac.2015.10.019]
21. Kanchi S, Sabela MI, Mdluli PS, Inamuddin, Bisetty K. Smartphone based bioanalytical and diagnosis applications: A review. Biosens Bioelectron 2018 Apr 15;102:136-149. [doi: 10.1016/j.bios.2017.11.021] [Medline: 29128716]
22. Baxter C, Carroll J, Keogh B, Vandelanotte C. Assessment of Mobile Health Apps Using Built-In Smartphone Sensors for Diagnosis and Treatment: Systematic Survey of Apps Listed in International Curated Health App Libraries. JMIR Mhealth Uhealth 2020 Feb 03;8(2):e16741 [FREE Full text] [doi: 10.2196/16741] [Medline: 32012102]
23. Vashist SK, Schneider EM, Luong JHT. Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management. Diagnostics (Basel) 2014 Aug 18;4(3):104-128 [FREE Full text] [doi: 10.3390/diagnostics4030104] [Medline: 26852680]

24. Al Ameen M, Liu J, Kwak K. Security and privacy issues in wireless sensor networks for healthcare applications. J Med Syst 2012 Feb;36(1):93-101 [FREE Full text] [doi: 10.1007/s10916-010-9449-4] [Medline: 20703745]
25. Pantelopoulos A, Bourbakis N. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Trans. Syst., Man, Cybern. C 2010 Jan;40(1):1-12. [doi: 10.1109/tsmcc.2009.2032660]
26. Majumder S, Mondal T, Deen MJ. Wearable Sensors for Remote Health Monitoring. Sensors (Basel) 2017 Jan 12;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
27. Deen MJ. Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers Ubiquit Comput 2015 Jun 26;19(3-4):573-599. [doi: 10.1007/s00779-015-0856-x]
28. Kpokiri EE, Marley G, Tang W, Fongwen N, Wu D, Berendes S, et al. Diagnostic Infectious Diseases Testing Outside Clinics: A Global Systematic Review and Meta-analysis. Open Forum Infect Dis 2020 Oct;7(10):ofaa360 [FREE Full text] [doi: 10.1093/ofid/ofaa360] [Medline: 33072806]
29. Hernández-Neuta I, Neumann F, Brightmeyer J, Ba Tis T, Madaboosi N, Wei Q, et al. Smartphone-based clinical diagnostics: towards democratization of evidence-based health care. J Intern Med 2019 Jan;285(1):19-39 [FREE Full text] [doi: 10.1111/joim.12820] [Medline: 30079527]
30. Ding X, Mauk MG, Yin K, Kadimisetty K, Liu C. Interfacing Pathogen Detection with Smartphones for Point-of-Care Applications. Anal Chem 2019 Jan 02;91(1):655-672 [FREE Full text] [doi: 10.1021/acs.analchem.8b04973] [Medline: 30428666]
31. Xu D, Huang X, Guo J, Ma X. Automatic smartphone-based microfluidic biosensor system at the point of care. Biosens Bioelectron 2018 Jul 01;110:78-88. [doi: 10.1016/j.bios.2018.03.018] [Medline: 29602034]
32. You DJ, Park TS, Yoon J. Cell-phone-based measurement of TSH using Mie scatter optimized lateral flow assays. Biosens Bioelectron 2013 Feb 15;40(1):180-185. [doi: 10.1016/j.bios.2012.07.014] [Medline: 22863118]
33. Herrera S, Vallejo AF, Quintero JP, Arévalo-Herrera M, Cancino M, Ferro S. Field evaluation of an automated RDT reader and data management device for Plasmodium falciparum/Plasmodium vivax malaria in endemic areas of Colombia. Malar J 2014 Mar 10;13:87 [FREE Full text] [doi: 10.1186/1475-2875-13-87] [Medline: 24612585]
34. Shekalaghe S, Cancino M, Mavere C, Juma O, Mohammed A, Abdulla S, et al. Clinical performance of an automated reader in interpreting malaria rapid diagnostic tests in Tanzania. Malar J 2013 Apr 24;12:141 [FREE Full text] [doi: 10.1186/1475-2875-12-141] [Medline: 23617722]
35. Kanakasabapathy MK, Sadasivam M, Singh A, Preston C, Thirumalaraju P, Venkataraman M, et al. An automated smartphone-based diagnostic assay for point-of-care semen analysis. Sci Transl Med 2017 Mar 22;9(382):eaai7863 [FREE Full text] [doi: 10.1126/scitranslmed.aai7863] [Medline: 28330865]
36. Allan-Blitz L, Vargas SK, Konda KA, Herbst de Cortina S, Cáceres CF, Klausner JD. Field evaluation of a smartphone-based electronic reader of rapid dual HIV and syphilis point-of-care immunoassays. Sex Transm Infect 2018 Dec;94(8):589-593. [doi: 10.1136/sextrans-2017-053511] [Medline: 30126946]
37. Herbst de Cortina S, Bristow CC, Humphries R, Vargas SK, Konda KA, Caceres CF, et al. Laboratory Evaluation of a Smartphone-Based Electronic Reader of Rapid Dual Point-of-Care Tests for Antibodies to Human Immunodeficiency Virus and Treponema pallidum Infections. Sex Transm Dis 2017 Jul;44(7):412-416 [FREE Full text] [doi: 10.1097/OLQ.0000000000000628] [Medline: 28604483]
38. Hu H, Liu X, Zhang Z, Xu X, Shi L, Fu G, et al. Increasing HIV Incidence among Men Who Have Sex with Men in Jiangsu Province, China: Results from Five Consecutive Surveys, 2011-2015. Int J Environ Res Public Health 2016 Aug 06;13(8):795 [FREE Full text] [doi: 10.3390/ijerph13080795] [Medline: 27509513]
39. Tang W, Tang S, Qin Y, Zhang Y, Zhang W, Liu C, et al. Will Gay Sex-Seeking Mobile Phone Applications Facilitate Group Sex? A Cross-Sectional Online Survey among Men Who Have Sex with Men in China. PLoS One 2016;11(11):e0167238 [FREE Full text] [doi: 10.1371/journal.pone.0167238] [Medline: 27880823]
40. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV Test Promotion Videos: A Noninferiority Randomized Controlled Trial in China. Clin Infect Dis 2016 Jun 01;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]
41. Marley G, Kang D, Wilson EC, Huang T, Qian Y, Li X, et al. Introducing rapid oral-fluid HIV testing among high risk populations in Shandong, China: feasibility and challenges. BMC Public Health 2014 May 03;14:422 [FREE Full text] [doi: 10.1186/1471-2458-14-422] [Medline: 24884431]
42. Andone I, Błaszkiewicz K, Eibes M, Trendafilov B, Montag C, Markowetz A. How age and gender affect smartphone usage. 2016 Sep 16 Presented at: UbiComp '16: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2016; Germany. [doi: 10.1145/2968219.2971451]
43. Veronese V, Clouse E, Wirtz AL, Thu KH, Naing S, Baral SD, et al. "We are not gays… don't tell me those things": engaging 'hidden' men who have sex with men and transgender women in HIV prevention in Myanmar. BMC Public Health 2019 Jan 14;19(1):63 [FREE Full text] [doi: 10.1186/s12889-018-6351-3] [Medline: 30642303]
44. Zhao Y, Zhu X, Pérez AE, Zhang W, Shi A, Zhang Z, et al. MHealth approach to promote Oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146 [FREE Full text] [doi: 10.1186/s12889-018-6046-9] [Medline: 30261856]

45. Balán IC, Lopez-Rios J, Nayak S, Lentz C, Arumugam S, Kutner B, et al. SMARTtest: A Smartphone App to Facilitate HIV and Syphilis Self- and Partner-Testing, Interpretation of Results, and Linkage to Care. AIDS Behav 2020 May;24(5):1560-1573 [FREE Full text] [doi: 10.1007/s10461-019-02718-y] [Medline: 31679075]
46. Lei Y, Zhang K, Xiao X, Tang C, Li X, Wang H. Sexual Roles, Risk Sexual Behaviours, and HIV Prevalence among Men who Have Sex with Men Seeking HIV Testing in Changsha, China. Curr HIV Res 2018;16(2):174-181 [FREE Full text] [doi: 10.2174/1570162X16666180711101643] [Medline: 29992887]
47. Lorente N, Champenois K, Blanche J, Préau M, Suzan-Monti M, Mora M, et al. Investigating Recent Testing among MSM: Results from Community-Based HIV Rapid Testing Attendees in France. J Sex Transm Dis 2013;2013:648791 [FREE Full text] [doi: 10.1155/2013/648791] [Medline: 26316962]
48. Yan H, Yang H, Raymond HF, Li J, Shi L, Huan X, et al. Experiences and correlates of HIV self-testing among men who have sex with men in Jiangsu province, China. AIDS Behav 2015 Mar;19(3):485-491 [FREE Full text] [doi: 10.1007/s10461-014-0968-8] [Medline: 25480598]
49. Njau B, Lisasi E, Damian DJ, Mushi DL, Boulle A, Mathews C. Feasibility of an HIV self-testing intervention: a formative qualitative study among individuals, community leaders, and HIV testing experts in northern Tanzania. BMC Public Health 2020 Apr 15;20(1):490 [FREE Full text] [doi: 10.1186/s12889-020-08651-3] [Medline: 32293370]
50. Njau B, Covin C, Lisasi E, Damian D, Mushi D, Boulle A, et al. A systematic review of qualitative evidence on factors enabling and deterring uptake of HIV self-testing in Africa. BMC Public Health 2019 Oct 15;19(1):1289 [FREE Full text] [doi: 10.1186/s12889-019-7685-1] [Medline: 31615461]
51. Xun H, Kang D, Huang T, Qian Y, Li X, Wilson EC, et al. Factors associated with willingness to accept oral fluid HIV rapid testing among most-at-risk populations in China. PLoS One 2013;8(11):e80594 [FREE Full text] [doi: 10.1371/journal.pone.0080594] [Medline: 24260432]
52. Han L, Wei C, Muessig KE, Bien CH, Meng G, Emch ME, et al. HIV test uptake among MSM in China: Implications for enhanced HIV test promotion campaigns among key populations. Glob Public Health 2017 Jan;12(1):31-44 [FREE Full text] [doi: 10.1080/17441692.2015.1134612] [Medline: 26785328]
53. Tucker JD, Wong FY, Nehl EJ, Zhang F. HIV testing and care systems focused on sexually transmitted HIV in China. Sex Transm Infect 2012 Mar;88(2):116-119 [FREE Full text] [doi: 10.1136/sextrans-2011-050135] [Medline: 22345024]

Abbreviations

AOR: adjusted odds ratio
CDC: Centers for Disease Control and Prevention
HIVST: HIV self-testing
MSM: men who have sex with men
SER: smartphone-based electronic reader

Edited by R Kukafka; submitted 13.12.20; peer-reviewed by X Jin, J Jiang, C Lyons; comments to author 15.03.21; revised version received 21.04.21; accepted 08.10.21; published 19.11.21.

Please cite as:
Marley G, Fu G, Zhang Y, Li J, Tucker JD, Tang W, Yu R
Willingness of Chinese Men Who Have Sex With Men to Use Smartphone-Based Electronic Readers for HIV Self-testing: Web-Based Cross-sectional Study
J Med Internet Res 2021;23(11):e26480
URL: https://www.jmir.org/2021/11/e26480
doi:10.2196/26480
PMID:34806988

Original Paper

Successes of and Lessons From the First Joint eHealth Program of the Dutch University Hospitals: Evaluation Study

Anneloek Rauwerdink[1], MD; Marise J Kasteleyn[2,3], PhD; Niels H Chavannes[2,3], MD, PhD; Marlies P Schijven[4,5], MD, MHS, PhD

[1]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Amsterdam, Netherlands

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

[3]National eHealth Living Lab, Leiden, Netherlands

[4]Department of Surgery, Amsterdam Gastroenterology and Metabolism, Amsterdam UMC, Amsterdam, Netherlands

[5]Citrien Fund program eHealth, Amsterdam, Netherlands

Corresponding Author:
Anneloek Rauwerdink, MD
Department of Radiology and Nuclear Medicine
Amsterdam UMC
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 20 566 9111
Email: anneloek@hotmail.com

Abstract

Background: A total of 8 Dutch university hospitals are at the forefront of contributing meaningfully to a future-proof health care system. To stimulate nationwide collaboration and knowledge-sharing on the topic of evidence-based eHealth, the Dutch university hospitals joined forces from 2016 to 2019 with the first *Citrien Fund (CF) program eHealth*; 29 eHealth projects with various subjects and themes were selected, supported, and evaluated. To determine the accomplishment of the 10 *deliverables* for the *CF program eHealth* and to contribute to the theory and practice of formative evaluation of eHealth in general, a comprehensive evaluation was deemed essential.

Objective: The first aim of this study is to evaluate whether the 10 deliverables of the *CF program eHealth* were accomplished. The second aim is to evaluate the progress of the 29 eHealth projects to determine the barriers to and facilitators of the development of the *CF program eHealth* projects.

Methods: To achieve the first aim of this study, an evaluation study was carried out using an adapted version of the Commonwealth Scientific and Industrial Research Organization framework. A mixed methods study, consisting of a 2-part questionnaire and semistructured interviews, was conducted to analyze the second aim of the study.

Results: The 10 deliverables of the *CF program eHealth* were successfully achieved. The program yielded 22 tangible eHealth solutions, and significant knowledge on the development and use of eHealth solutions. We have learned that the patient is enthusiastic about accessing and downloading their own medical data but the physicians are more cautious. It was not always possible to implement the Dutch set of standards for interoperability, owing to a lack of information technology (IT) capacities. In addition, more attention needed to be paid to patients with low eHealth skills, and education in such cases is important. The eHealth projects' progress aspects such as *planning*, *IT services*, and *legal* played an important role in the success of the 29 projects. The in-depth interviews illustrated that a novel eHealth solution should fulfill a need, that partners already having the knowledge and means to accelerate development should be involved, that clear communication with IT developers and other stakeholders is crucial, and that having a dedicated project leader with sufficient time is of utmost importance for the success of a project.

Conclusions: The 8 Dutch university hospitals were able to collaborate successfully and stimulate through a bottom-up approach, nationwide eHealth development and knowledge-sharing. In total, 22 tangible eHealth solutions were developed, and significant eHealth knowledge about their development and use was shared. The eHealth projects' progress aspects such as *planning*, *IT services*, and *legal* played an important role in the successful progress of the projects and should therefore be closely monitored when developing novel eHealth solutions.

International Registered Report Identifier (IRRID): RR2-10.1016/j.ceh.2020.12.002

(J Med Internet Res 2021;23(11):e25170) doi:10.2196/25170

KEYWORDS

CSIRO framework; evaluation strategy; eHealth; telemedicine; qualitative research; formative evaluation; digital health

Introduction

Background

The global population is increasing rapidly, and the number of people aged ≥60 years is expected to double by 2050 [1,2]. The direct consequences of global aging include rising health care expenditures and a potential shortage of health care professionals. The current COVID-19 pandemic has further uncovered the vulnerability of our health care systems and accessibility of, for instance, our hospitals in times of social distancing and a shortage of capacity [3,4].

Therefore, governments are increasingly debating which health care reforms are necessary to preserve the quality of our health care system and how to effectively deliver care from a distance. The concept of *eHealth* may support the necessary reforms [5]. By implementing eHealth solutions in daily practice, it is expected that health care processes would be executed more efficiently and subsequently time and costs would be saved [6-8]. Moreover, the use of eHealth can increase patient participation and empowerment [9,10]. In 2015, the World Health Organization Global Observatory for eHealth explored eHealth developments and investigated how eHealth can support universal health coverage [11]. The report considered eHealth foundation built through policy development, funding approaches, and training of students and professionals. Policy development is of utmost importance to counteract the *fragmentation* of eHealth initiation [12]. In addition to policy development, it is important to investigate how to develop and implement novel eHealth solutions at scale successfully. Schreiweis et al [13] summarized the critical factors influencing the implementation and adoption of eHealth. They described the perceived barriers, such as added workload, problems with financing, and missing fit in organizational structures. In addition, a lack of system interoperability is a well-known and frustrating issue in preventing the sustainable implementation of eHealth [14]. Finally, it seems that developers, evaluators, and physicians find it difficult to learn from successful initiatives that come from external sources, such as those from other disciplines or from outside the region [15]. They have the so-called “not-invented-here syndrome.”

In the Netherlands, as part of the Dutch national eHealth strategy, the *Citrien Fund* (CF) was established for the period of 2014 to 2018 by the government-funded Netherlands Organisation for Health Research and Development (in Dutch: ZonMw). The CF aims to contribute to a sustainable health care system by stimulating collaboration between 8 Dutch university hospitals, and between the university hospitals and other health care organizations. The CF supports 5 programs with different themes [16]. This study focuses on the *CF program eHealth*, which took place from 2016 until the beginning of 2019. This first nationwide university hospital eHealth collaboration mainly focused on the constitution of a strong collaborative framework to discuss present eHealth issues, on the development of a wide array of novel eHealth solutions—with the most successful being scaled in a subsequent program—and on sharing of eHealth knowledge. The overall aim of the *CF program eHealth* was to accomplish 11 predefined deliverables (Textbox 1), which were drafted upon knowledge gaps within the Dutch eHealth landscape. To achieve deliverables 2 to 11, 29 eHealth research projects were conducted.

Objectives

By conducting a comprehensive evaluation of the *CF program eHealth*, including determination of the barriers to and facilitators of the development of each of the 29 projects, this study aims to assess the successes of and lessons from the *CF program eHealth*. This study also aims to reduce the scarcity of formative eHealth evaluations and to become a useful case study for the eHealth evaluators of eHealth development programs [17-19]. Subsequently, the successful future development and implementation of eHealth, in general, might be enhanced. This study consists of 2 aims: (1) to evaluate whether the 10 deliverables of *CF program eHealth* were accomplished and (2) to evaluate the progress of the 29 eHealth projects to determine the barriers to and facilitators of the development of the *CF program eHealth* projects.

Textbox 1. Deliverables of the Citrien Fund program eHealth [20].

1. One coordinating Dutch Federation of University Medical Centers vision on eHealth and eHealth road map (the accomplishment of this deliverable was reported in a previous study and is, therefore, not part of this study [21]).
2. A virtual nationwide expertise center for eHealth.
3. International positioning by promoting in journals and media and during the closure event.
4. Conditions for downloading medical data described and, if possible, realized.
5. Blueprint for interoperability between hospital information systems and electronic health records.
6. Agreements and standards for data sharing between consumer and professional eHealth.
7. A framework for regional collaboration for effective implementation of eHealth.
8. Models that can strengthen the empowerment of the patient.
9. A developed multidisciplinary infrastructure to stimulate the development of digital health.
10. Development, evaluation, and implementation of eHealth instruments in collaboration with companies and start-ups.
11. A blueprint for education in eHealth competencies and skills for health care professionals.

Methods

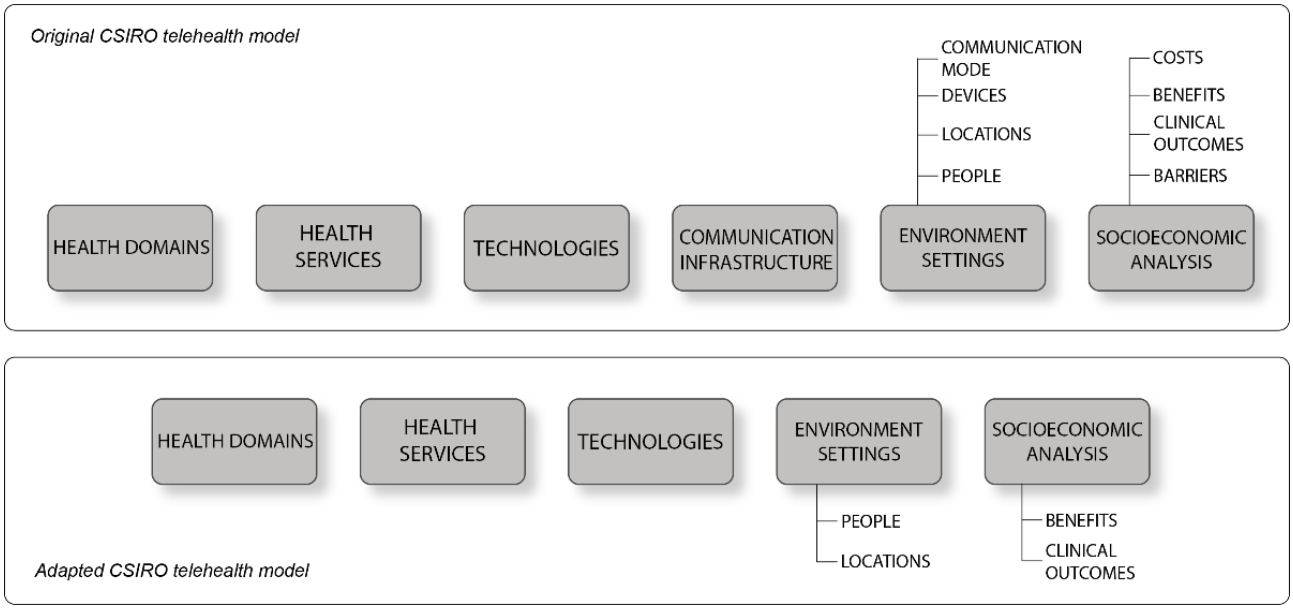
Setting

As described in the previous study protocol, each of the 8 Dutch university hospitals was asked to propose 3 or 4 eHealth research projects to be carried out in their own university hospital. Although the university hospitals had a carte blanche for the projects they proposed, every project had to contribute to one or more of the 10 program deliverables. In total, 29 projects covering a wide array of eHealth themes were carried out from June 2016 to January 2019 [16]. The projects delivered either a tangible eHealth solution or provided knowledge about the use or development of an eHealth solution. Each project was managed by a dedicated local project leader, with 29 project leaders in total. At the initiation of the *CF program eHealth*, each project leader drafted a detailed project plan describing which of the 10 program deliverables the project aimed to contribute, the project's objectives, and a detailed timeline. The project plan was approved by the supervising steering committee, which consisted of 2 representatives of each university hospital. In a 3-month meeting, the steering committee monitored the progress of the projects and advised the project leaders. In addition, at the midterm (December 2017) and end-term (October 2018) of the program, the project leaders presented their projects' progress to the steering committee, and a majority decision regarding the continuation of the project was made.

Evaluation Study

An adapted version of the *Commonwealth Scientific and Industrial Research Organization (CSIRO) framework for evaluating telehealth trials or programs* (Figure 1) was deemed the most suitable framework for the evaluation of the *CF program eHealth* (Textbox 1), as described in a previous study protocol paper [16,22]. The following aspects of the adapted CSIRO framework for each project were described: health domain, health service, technology, environment setting, and (clinical) outcomes or evident benefits. Owing to the wide array of projects, covering various topics and using diverse study designs, the last aspect was broadly defined, ranging from usability outcomes to clinical outcomes, and qualitative eHealth insights. The adapted CSIRO framework was incorporated into the *Citrien Fund – mapping table* (Multimedia Appendix 1), which also presents general information and the completion status for each project's deliverables.

Figure 1. The original Commonwealth Scientific and Industrial Research Organization (CSIRO) model versus the adapted CSIRO telehealth model.



In the *program deliverables* column of the mapping table, the deliverables that the project aimed to achieve were indicated with red, orange, or green dots depending on the level of accomplishment at end-term. The red dot indicated that the project failed to accomplish the deliverable, orange indicated partial completion, and green indicated full accomplishment of the deliverable. Based on this overview, a short summary for the completion of each deliverable is provided.

During the end-term presentation, the project leaders presented the project's main findings to the steering committee. This presentation included all aspects of the adapted CSIRO framework and was used to systematically collect the data for each project.

Mixed Methods Study

Through a mixed methods approach, consisting of a questionnaire and a semistructured interview, the barriers to and facilitators of the development of the 29 eHealth projects *eHealth* were determined. Previously, it had been found that 7 eHealth project progress aspects were useful for monitoring the progress of eHealth projects: planning, needs assessment, policy or organization, technology, ethics, legal, and finance (Figure 2) [16]. By quantitatively and qualitatively evaluating these 7 aspects for the 29 projects, we aimed to obtain insights into which barriers and facilitators were important for the successful development of the *CF program eHealth* projects.

Figure 2. eHealth project progress aspects.

Questionnaire

All project leaders were asked to fill out a self-developed 2-part questionnaire. The main concepts that the questionnaire covered were the eHealth project progress aspects: planning, policy or organization, technology, ethics, legal, and finance. Various question formats were used, including yes or no, multiple-choice, and 4-point Likert scale questions. Part one had to be completed at the midterm (52 questions) and part two at the end-term (55 questions) for each project. With this longitudinal aspect, we aimed to evaluate the change in magnitude of the eHealth project progress over time. The topics included respondent demographics and items related to eHealth project progress aspects. The questionnaire was completed using the web-based survey software SurveyMonkey (SurveyMonkey, Inc) [23]. The respondents received an email with an invitation link to the questionnaire and reminder emails were sent to them at 3- and 5-week intervals.

Interview

At the end of the *CF program eHealth*, one project leader from each university hospital was randomly selected by drawing lots and interviewed by the coordinating researcher (AR) for a more in-depth exploration of the role of the eHealth project progress aspects in their project. The semistructured interviews were held by telephone, following a previously composed interview guide (Multimedia Appendix 2), which had a subset of open questions on each eHealth project progress aspect and used an iterative approach. The same set of questions was used in all interviews and, if necessary, adjusted along the way. The interview guide was composed of inputs from the results of the questionnaire and notes from the 3 monthly meetings of the steering committee. It was estimated that saturation was reached after 8 interviews; however, if saturation had not been reached, more project leaders were interviewed.

Data Analysis

The quantitative questionnaire data were analyzed by calculating descriptive statistics using Microsoft Excel. Continuous data were summarized using median and IQR. Categorical data were presented as frequency counts with percentages.

The 8 interviews were digitally recorded with the permission of the project leaders and transcribed verbatim. AR analyzed the transcripts according to the 6-step thematic analysis framework of Braun and Clarke [24] and discussed the results with the last author (MPS). First, we read the transcripts thoroughly to familiarize ourselves with the data. Second, notes were placed next to the text to generate the initial codes. Finally, major themes were identified and defined, and the data were further coded and sorted into themes and subthemes. The data from each theme were summarized into descriptions concerning the contribution of the 7 eHealth project progress aspects in the successful performance of an eHealth project.

Results

The results of the evaluation study are schematically presented in the *Citrien Fund - mapping table* (Multimedia Appendix 3). In total, 22 projects developed a tangible eHealth solution, and 7 projects acquired knowledge about the development and use of eHealth solutions. The projects were conducted in various *health domains*, of which internal medicine represented the

biggest share. In the *health services* column table, it is shown that most of the projects contributed to nonclinical services, such as education, administration, and research. Projects contributing to clinical services, such as treatment, monitoring, and diagnostics, were mentioned in a minority of cases. In the *environment setting*, 17 projects focused on patients, 14 on health care providers or physicians and 6 on researchers. A total of 14 projects were related to a home-based environment and 6 to a hospital environment.

Evaluation Study

All 10 deliverables of the *CF program eHealth* were successfully achieved. The contributions of the 29 projects to the 10 deliverables are summarized in Table 1. In summary, most of the projects (19/29, 66%) contributed to models that strengthen the patients' directing role (deliverable 8), making *strengthening the patients' directing role* the most prominent theme of the *CF program eHealth*. We learned that the patients were enthusiastic about accessing and downloading their own medical data on the web (deliverable 4), but physicians were more cautious, and it was not always possible to implement the Dutch set of standards for interoperability owing to a lack of information technology (IT) capacities (deliverable 5). We also learned that considerable attention should be paid to eHealth literacy when developing novel eHealth solutions (deliverable 4). Furthermore, the establishment of alternative communication infrastructures (deliverable 9) between the caregiver within the hospital and the patient outside the hospital was investigated and considered very relevant.

Table 1. Summary of major findings per deliverable.

Deliverable	Projects[a]	Conclusion
2	24	A Dutch nationwide web-based expertise center for eHealth was established, containing an eHealth toolkit with the various delivered products and the acquired knowledge [25].
3	4, 13, 15, 16, 20, and 21	Various projects published their results in national and international (scientific) journals [26-33]. In addition, an e-book has been published [34].
4	5, 9, 10, 13, 19, 23, 24, and 26	In general, the patients were enthusiastic about accessing their medical data on the web and downloading them. Physicians were still holding back. eHealth literacy must be considered in the development and implementation of eHealth. One of the university hospitals dealt with information exchange between the primary care and the hospital, for which a legal framework had been set.
5	5, 13, 19, 23, and 25	It was not always possible to implement the Dutch set of standards for interoperability. Integrating different IT[b] systems for exchanging data was complex and might not be desirable in a pilot phase.
6	14, 19, 23, 25, and 26	An important insight obtained was that when exchanging data, the skills of the end user should be considered. Attention should also be paid to patients with low eHealth skills.
7	5, 16, 18, 23, 24, and 29	Within the projects, there was frequent co-operation with IT developers in the region and the first- and second-line health care institutions.
8	1, 4, 5, 6, 7, 10, 13, 14, 15, 16, 17, 18, 22, 24, 25, 26, 27, 28, and 29	Most of the projects (19/29, 66%) contributed to models that strengthen the patients' directing role. For example, 2 e-learnings were developed, in which both the patient and the caregiver received the tools to make better decisions together. In addition, several mobile apps or web-based applications that were developed, for example a medical dashboard, a patient coach, an app for glycemic index, a web-based blog and forum for patients with Alzheimer, and a home-based blood pressure monitor for high-risk pregnant women, reinforced the patients' directing role.
9	15, 16, 18, 19, and 28	Several projects focused on establishing alternative communication infrastructures between a hospital and a patient outside the hospital. In addition, for the development of these new type of eHealth solutions, the projects required close co-operation and consultation with researchers, patients, informal caregivers, IT services, and lawyers.
10	1, 2, 3, 4, 6, 7, 8, 11, 13, 14, 15, 16, 17, 20, 21, 24, 26, 27, 28, and 29	A website with a forum and a blog was developed for patients with Alzheimer. Also, a total of 8 mobile apps were developed. Various wearables were tested for the home monitoring of patients. A clinical data science eBook was made, and several e-learnings were developed. The efficacy and effectiveness of the various eHealth solutions were scientifically evaluated.
11	3, 8, 12, 14, 15, 17, 21, 24, and 27	Several projects contributed to improving eHealth education. For example, the *develop your own eHealth app* project developed an education module in which medical students learned about eHealth and the necessary eHealth skills. Another project translated the English language Apple Research Kit into a Dutch variant and offered researchers a guide on how to conduct eHealth research with the kit. Finally, some projects also focused on patient education and how to enable patients with low health skills to work with eHealth tools.

[a]The project numbers correspond with the projects illustrated in the Citrien Fund - mapping table in Multimedia Appendix 3.

[b]IT: information technology.

Owing to insufficient progress, one project was prematurely terminated by the steering committee after the midterm evaluation, and another 3 projects were prematurely terminated after the end-term evaluation. These projects are indicated with an asterisk (*) in the *Citrien Fund - mapping table*. As shown in the *(Clinical) outcomes or evident benefits* column, in 2 of the 4 cases, personnel matters were responsible for the insufficient progress, in one case there was an issue with the

intraoperability of the proposed eHealth solution with the existing IT system and in the last case there was an *IT freeze* hospital-wide because of the implementation of a new electronic health record system.

Mixed Methods Study

Questionnaire

The 2-part questionnaire was completed by all the participating project leaders in November 2017 and October 2018. The first part was completed by 29 project leaders, and the second part was completed by 27 project leaders owing to the premature termination of 1 project and the early completion of another. Tables 2 and 3 show the main characteristics of the participants and the projects, respectively.

Table 2. Demographics of project leaders (N=29)[a].

Demographics	Values, n (%)
Gender	
Female	17 (59)
Male	12 (41)
Degree[b]	
Medicine	7 (19)
Psychology	4 (11)
Health sciences	3 (8)
(Medical) biology	2 (6)
Communication	2 (6)
Other[c]	18 (50)
Employment	
Hospital	22 (76)
Outside hospital	3 (10)
Both	4 (14)
Weekly time spent at the project (hours)	
<5	9 (31)
5-10	12 (41)
10-15	3 (10)
>15	5 (17)
Age (years), median (IQR)	33 (29-44.5)

[a]Measured midterm.

[b]More degrees per project leader possible.

[c]One degree per other specialty.

Table 3. General characteristics of projects (N=29)[a].

Characteristics	Values
Reasons for participating in the *CF*[b]*program eHealth, n (%)*	
Subsidy	25 (86)
Publicity	19 (66)
Collaboration	20 (69)
Other reasons	5 (17)
If the project was not accepted in *CF*[b]*program eHealth*, then, n (%)	
It would have remained a project plan	8 (28)
I would have actively searched for other means	21 (72)
Other existing means were allocated to the project	0 (0)
Number of people involved internally, n (%)	
0	1 (3)
1-2	3 (10)
3-4	15 (52)
>5	10 (34)
Number of people involved externally, n (%)	
0	6 (21)
1-2	11 (38)
3-4	6 (21)
>5	6 (21)
Number of monthly meetings with steering committee member, n (%)	
0-1	13 (45)
1-2	7 (24)
>2	4 (14)
Never	5 (17)
Time to acceptation in months, median (IQR)	0 (0-4)

[a]Measured midterm.

[b]CF: Citrien Fund.

The main reason for participating in the *CF program eHealth* was to receive funding (25/29, 86%). However, publicity (19/29, 66%) and collaboration (20/29, 69%) were also important reasons. Of the 29 projects, 25 (86%) had ≥3 people internally involved, and in 23 (79%) projects, people from outside the hospital, such as general practitioners, patients, and software developers, were involved as well.

The main results of the questions concerning *planning* have been presented in Table 4. At midterm, all the project leaders (29/29, 100%) estimated that they would be able to complete the selected program deliverables, as described in their project plan. However, at end-term, 4 project leaders (4/27, 15%) indicated that they might not be able to contribute to the deliverables. At midterm, almost half (12/29, 41%) of the project leaders indicated that their project planning was no longer up to date. Moreover, at the end of the study, the planning of the majority (23/27, 85%) was not up to date. In addition, at midterm, 6 project leaders (6/29, 21%) indicated that the time available for successful progress in the project was not sufficient; this share doubled (12/27, 44%) at the end of the questionnaire.

Table 4. eHealth project progress aspect *planning*.

Question	Midterm (n=29)		End-term (n=27)	
	Yes, n (%)	No, n (%)	Yes, n (%)	No, n (%)
Is the available time sufficient for successful progress of your project?	23 (79)	6 (21)	15 (56)	12 (44)
Is your project planning, as described in the project plan, still up to date?	17 (59)	12 (41)	4 (15)	23 (85)
As the project is progressing now, I expect to achieve the measurable goals as described in the project plan	25 (86)	4 (14)	19 (70)	8 (30)
I do not foresee any problems in contributing to program deliverables at the end of the Citrien Fund program eHealth, as described in my project plan	29 (100)	0 (0)	23 (85)	4 (15)

Table 5 presents results regarding the eHealth project progress aspects, *policy* or *organization*, *technology*, *ethics*, and *legal*. None of the topics was spared of inconvenience, with IT services and privacy issues representing the greatest shares in moderate to significant inconvenience.

Regarding the aspect of *finance*, 8 projects (8/29, 28%) received additional funding other than that from the *CF program eHealth* at the initiation of the project. Furthermore, at midterm, 8 project leaders (8/29, 28%) thought that they would need extra funding for the successful completion of their project. However, at the end of the study, 13 project leaders (13/27, 48%) stated that extra funding would be necessary for project completion. The funding received from the *CF program eHealth* was mostly insufficient to cover the personnel expenses and the implementation aims. After termination of the *CF program eHealth*, of the 27 projects, 13 (48%) still needed to find financial means to continue, while 4 (15%) already had the means and 10 (37%) did not need any.

Table 5. Inconvenience issues encountered during project execution (N=27).

	No inconvenience, n (%)	Some inconvenience, n (%)	Moderate inconvenience, n (%)	Significant inconvenience, n (%)	Not applicable, n (%)
Realizing contracts with third parties	7 (25)	6 (22)	7 (25)	3 (11)	4 (14)
Privacy issues, such as patient data protection	9 (33)	4 (14)	7 (25)	5 (18)	2 (7)
Review of medical ethics committee	7 (25)	8 (29)	5 (18)	4 (14)	3 (11)
Resistance from within the organization	8 (29)	9 (33)	8 (29)	2 (7)	0 (0)
Resistance from outside the organization	13 (48)	10 (37)	4 (14)	0 (0)	0 (0)
Information technology developers and support	7 (25)	5 (18)	6 (22)	6 (22)	3 (11)
Electronic health record supplier	8 (29)	3 (11)	0 (0)	3 (11)	13 (48)

Interview

In total, 8 project leaders were interviewed, representing the entire group of project leaders. After 7 interviews, saturation of information was reached. Three major themes, with subthemes, were identified (Textbox 2). The main findings of the interviews have been discussed under these 3 themes, supported by quotes (Multimedia Appendix 4) as examples of the participants' responses.

Textbox 2. Themes and subthemes identified from interviews.

Success factors for eHealth development and implementation

- Fulfill a need
- Outsource
- Communication
- Personnel

Essential third parties

- Information technology services
- Medical Ethical Committee
- Legislation

Flexibility

- Project planning
- Conducting research
- Effectiveness testing

Theme: Success Factors for eHealth Development and Implementation

While carrying out their eHealth projects, the project leaders encountered several relevant aspects that contributed to the success of their projects. In such cases, there should be an evident *need to fulfill*, or problem to solve, when developing an eHealth solution, although conducting a needs assessment was not deemed necessary. Regarding *outsourcing*, it was crucial to find the right (commercial) partners that already had the knowledge and means to accelerate development. Clear *communication* with IT developers and other stakeholders about the development and other concerns, for instance, estimated changes in routine care, was essential. Successful development of an eHealth solution may depend on the availability and dedication of the *personnel*, including the project leader.

Theme: Essential Third Parties

Owing to the immaturity and complexity of eHealth, several important topics related to its development and implementation have never been discussed before. Therefore, communication and close collaboration with third parties was of utmost importance.

Regarding *IT services*, the project leaders indicated that communication was the number one pitfall in successful collaboration. It should also be emphasized that IT development costs time, capacity, and money. *Medical ethical committees* may find it difficult to take a position because of the unknown impact or burden of a novel eHealth solution. Moreover, one should be informed about the impact of *legislation* on their project. In addition, a privacy impact assessment is often obligatory, which may cause a delay in planning.

Theme: Flexibility

eHealth solutions are considered complex interventions with multiple interacting components. This required some level of flexibility when it came down to project planning, conducting research, and effectiveness testing. *Project planning* is important in the initial stages. However, it should be possible in case of incidents to make timely adjustments.

Regarding *conducting research*, there were varying responses to whether it was challenging to find study participants to evaluate the eHealth project. However, it seems that when an eHealth solution can solve a relevant problem, patients are willing to participate. Furthermore, the study end point was difficult to determine because of the novel character of eHealth and the resulting lack of literature. eHealth solutions should be evaluated in a study context where possible. Nevertheless, it was considered unnecessary to conduct a randomized controlled trial to prove *effectiveness*, because, for example, patients had already experienced significant benefits while using a novel eHealth solution. In addition, while clinical effectiveness may be obligatory, it does not tell anything about the effectiveness of the eHealth solution.

Discussion

Principal Findings

Evaluation Study

The evaluation study targeted the first study aim. By systematically evaluating the accomplishment of 10 program deliverables, we were able to determine the successes and lessons of the *CF program eHealth*. In this study, we learned that patients are more enthusiastic about downloading their own medical data than physicians, that the lack of IT capacities plays a negative role in implementing the Dutch set of standards for interoperability, that considerable attention should be paid to patients with lower eHealth skills, and that establishing alternative communication infrastructures between caregivers and patients should be considered very important.

The 29 different eHealth research projects delivered 22 tangible eHealth solutions and significant knowledge about the development and use of eHealth solutions. Strengthening the

patient's directing role was the most prominent theme of the *CF program eHealth.*

Through the formation of a new collaborative network, the *CF program eHealth* was able to bring the 8 Dutch university hospitals together and therewith significantly improve the fragmented eHealth landscape in the Netherlands.

Mixed Methods Study

The mixed methods study targeted the second study aim. The questionnaire helped us learn that the eHealth project progress aspects *planning*, *technology*, and *legal* played an important role in successful development of the 29 projects. However, a lack of time with the individual project leaders, priorities other than implementing eHealth in IT services, and the never discussed before privacy issues, together with the relatively short *CF program eHealth* duration, caused project delays.

In the in-depth interviews, the themes: *success factors for eHealth development and implementation*, *essential third parties*, and *flexibility*, were identified as the 3 most important themes to pay attention to when carrying out an eHealth project.

National eHealth Programs

Although it seemed that the Netherlands was among the first countries to carry out a publicly funded university hospital eHealth program, there are other nationwide eHealth programs and initiatives with which the *CF program eHealth* could be compared.

The National Health Service (NHS), the publicly funded health care system in England, holds an active and well-organized digital section, the *NHS digital* [35]. One of its programs encompasses the initiation of a physical and conceptual digital, research, informatics, and virtual environments (DRIVE) unit, which explores, among other topics, how to gain insights from machine learning and artificial intelligence (AI). To the best of our knowledge, the program is not scientifically evaluated nor is it visible to the public what has been done. In addition, the NHS only collaborates with the Great Ormond Street Hospital, whereas the collaborative network of the *CF program eHealth* encompasses all university hospitals of the Netherlands. Furthermore, as described by Astana et al [12], the top-down approach of the NHS and fragmentation with respect to the organization and delivery of care, makes it a complex landscape for eHealth companies seeking to enter the system and scale up innovation [12]. In contrast, the *CF program eHealth* used a bottom-up approach to search for innovative initiatives, controlled by the 8 university hospitals rather than the government itself. Owing to the high visibility of this program on the web and the co-operation with stakeholders and patients, many useful insights into nationwide eHealth development were gathered and disseminated directly.

The Danish program *Patient@home* focused on rehabilitation and monitoring services to promote patient empowerment and support treatment at home [36]. In total, 30 projects were carried out through public-private collaborations between patients, research institutions, and other stakeholders. Although all the projects were developed using structured 5-phase innovation models, with the backgrounds and aims as described in detail on the program's website, no scientific results were reported on the website [37]. In the case of the *CF program eHealth*, projects may have benefited from a structured innovation model with subsequent phases of development and implementation. Especially in the case of the 4 projects that were prematurely terminated, an evaluation of the first phase of innovation—*need*—which includes technology screening, could have been beneficial. However, the well-thought program structure, consisting of an obligatory project plan and a planning stage at the initiation of a project and 2 project progress evaluations along the way, might have overcome the lack of a structured innovation model.

eHealth Project Progress Aspects

The mixed methods study found that the eHealth project progress aspects *planning*, *technology*, and *legal* were important aspects in relation to successful progress of the projects. From an implementation perspective, comparable results were described by Schreiweis et al [13]. After conducting a systematic literature and expert discussion analysis, the authors considered flexible funding, health outcomes, policies for using generated data for research, competition, and supporting laws and regulations, as important factors for success. Our study added insights into the successful project progress from a developmental perspective. For example, having a dedicated project leader was essential, as were flexible project planning, clear communication with IT services, and collaboration with (commercial) partners that already had the knowledge and means to accelerate development. Liu et al [38] studied the barriers to and facilitators of such an academia-industry collaboration. They identified the aspect timeline, consisting of longer time frames in research projects, contrasting with the greater emphasis on quick implementation in industry, as a barrier to successful academia-industry collaboration. To mitigate this, our study found that outlining and communicating openly about the goals and expectations may facilitate successful academia-industry collaborations.

Vedlūga and Mikulskienė [39] compiled a corpus of indicators to monitor the implementation of the national eHealth information system and proposed 5 key dimensions of stakeholder-driven performance elements for eHealth evaluation: human resources, financial resources, management resources, legal aspects, satisfaction with technological solutions, and design. These performance elements match with the results of our study, such as the issues caused by misunderstandings between clients and IT service providers, shortage of funding, and never discussed before legal issues. The element human resources was considered less important to eHealth development. However, close attention should be paid to the endemic problem of researchers performing work in their spare time.

Strengths and Limitations

One of the major strengths of the *CF program eHealth* was the open and collaborative organization with a bottom-up approach. For example, one representative from each of the 8 Dutch university hospitals took place in the steering committee and was responsible for monitoring and mentoring local projects. A solid foundation was laid for successful program progress and the achievement of program deliverables.

Another strength was the wide array of subjects and studies in the 29 eHealth projects. Although this variation may have reduced comparison possibilities and thus the external validity of projects, many valuable insights that proved important across settings were gained. However, by creating a systematic overview of the projects' findings in the evaluation study, their comparability was greatly enhanced. The combination of an evaluation study with a mixed methods study to evaluate the progress of eHealth projects in detail further strengthens our study findings.

A study limitation was that the *CF program eHealth* deliverables were not formulated through the well-known Specific, Measurable, Attainable, Relevant, and Timely formats. Therefore, it might be debatable whether the 10 deliverables were truly completed. Another limitation regarding the achievement of the deliverables might be the finding that most of the project leaders indicated in the questionnaire that their project planning was no longer up to date at the end of the *CF program eHealth*. However, a *not anymore up to date* project plan did not necessarily mean project failure and, therefore, its failure to achieve deliverables. Project planning was relevant to the initial direction and evaluation progress but on-time completion of project planning was not compulsory when drawing the final conclusions about the results of a project.

A limitation regarding the mixed methods study might have been the number of interviewed project leaders. However, saturation of information was accomplished after 7 interviews, and the interviewed project leaders were a representative sample of the whole group.

A final limitation was that we used a self-developed questionnaire to assess the eHealth project progress aspects of the 29 projects. We decided to develop a questionnaire owing to the relatively small group of participants and the limited time frame. Although valuable insights were gained, methodologically, it would have been stronger if some level of content validity and construct validity was carried out.

Future Perspectives

After termination of the *CF program eHealth*, the most successful projects have now scaled up nationwide in a subsequent edition of the *CF program eHealth*, focusing on implantation and upscaling [40]. The lack of sequential funding, as indicated by many in the Mixed Methods Study section of this paper, will be overcome for these projects. In addition, the assumption that most eHealth projects suffer from *pilotitis* (ie, projects will never pass the pilot phase), might also be overcome with the sequel of *CF program eHealth* [17].

The COVID-19 pandemic has accelerated the uptake and implementation of existing eHealth solutions [41,42]. Although care delivered at a distance has been already possible for many years, actual scaled-up use is still lacking. The aforementioned aspects of technology and law, which slowed the progress of the development of the projects, might be positively influenced and could take eHealth to a more mature level in the aftermath of this crisis. However, new points of discussion will need to be taken care of, such as how to take care of data privacy and legislation if a nationwide COVID-19 mobile app was implemented [43].

The *CF program eHealth* has proven that nationwide collaboration on the theme of eHealth between the 8 Dutch university hospitals and related commercial parties is possible and diminishes eHealth fragmentation. To truly preserve and improve the quality of our health care system in the light of global aging, we should strive for the elimination of eHealth fragmentation and national and international eHealth collaborations, such as the *CF program eHealth*, should be stimulated by governments and the European Union.

Conclusions

The 8 Dutch university hospitals were able to successfully collaborate and stimulate nationwide evidence-based eHealth development using a bottom-up approach. In total, 22 novel eHealth solutions with various subjects were developed, and significant knowledge about eHealth development and use was established. The aspects *planning*, *technology*, and *legal* played an important role in successful progress of the projects and should therefore be closely monitored while developing novel eHealth solutions or when implementing existing solutions. To further counteract eHealth fragmentation and take the next step from development to upscaling, a subsequent *CF program eHealth* will be carried out.

Acknowledgments

The authors would like to thank all the projects leaders of the CF program eHealth for investing their time and effort to participate in the mixed methods study.

Special thanks go to J Oegema for providing language help and writing assistance.

This study of the CF program eHealth was funded by the Netherlands Organization for Health Research and Development (ZonMw). ZonMw was not included in this study or in the article. The Dutch Federation of University Hospitals (in Dutch: Nederlandse Federatie van Universitair Medische Centra) is responsible for the execution of the CF, and ZonMw is responsible for providing governance and process monitoring.

We would like to thank all of the members of the steering committee of the Citrien fund program eHealth for their time and efforts spend on the program and on this study in particular. Steering committee members: Pim WJJ Assendelft, Erik Buskens, Niels H Chavannes, Peter Van Felius, Jan A Hazelzet, Vivian Hoogland, Marise J Kasteleyn, Rudolf B Kool, Pieter L Kubben, Maarten MH Lahr, Hyleco H Nauta, Marlies P Schijven, Wouter Van Solinge, Nynke Venema-Taat, Irma M Verdonck-de Leeuw and Erik Zwarter.

Authors' Contributions

All authors critically read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Empty Citrien Fund - mapping table.
[XLS File (Microsoft Excel File), 31 KB - jmir_v23i11e25170_app1.xls]

Multimedia Appendix 2
Interview guide.
[DOCX File , 14 KB - jmir_v23i11e25170_app2.docx]

Multimedia Appendix 3
Citrien Fund - mapping table.
[DOCX File , 32 KB - jmir_v23i11e25170_app3.docx]

Multimedia Appendix 4
Interview quotes.
[DOCX File , 22 KB - jmir_v23i11e25170_app4.docx]

References

1. World Populations Prospects. United Nations. 2017. URL: https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf [accessed 2019-01-25]
2. About global aging. Global Aging Institute. URL: http://www.globalaginginstitute.org/about-gai/about-global-aging.html [accessed 2019-08-25]
3. Rizzi AM, Polachek WS, Dulas M, Strelzow JA, Hynes KK. The new 'normal': rapid adoption of telemedicine in orthopaedics during the COVID-19 pandemic. Injury 2020 Dec;51(12):2816-2821 [FREE Full text] [doi: 10.1016/j.injury.2020.09.009] [Medline: 32951916]
4. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
5. 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]
6. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]
7. Freed J, Lowe C, Flodgren G, Binks R, Doughty K, Kolsi J. Telemedicine: is it really worth it? A perspective from evidence and experience. J Innov Health Inform 2018 Mar 15;25(1):14-18 [FREE Full text] [doi: 10.14236/jhi.v25i1.957] [Medline: 29717950]
8. Charles BL. Telemedicine can lower costs and improve access. Healthc Financ Manage 2000 Apr;54(4):66-69. [Medline: 10915354]
9. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
10. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
11. Global diffusion of eHealth: making universal health coverage achievable. World Health Organization. 2016. URL: http://apps.who.int/iris/bitstream/10665/252529/1/9789241511780-eng.pdf [accessed 2019-01-01]
12. Asthana S, Jones R, Sheaff R. Why does the NHS struggle to adopt eHealth innovations? A review of macro, meso and micro factors. BMC Health Serv Res 2019 Dec 21;19(1):984 [FREE Full text] [doi: 10.1186/s12913-019-4790-x] [Medline: 31864370]
13. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
14. Shull JG. Digital health and the state of interoperable electronic health records. JMIR Med Inform 2019 Nov 01;7(4):e12712 [FREE Full text] [doi: 10.2196/12712] [Medline: 31682583]

15. Myers CG, Sutcliffe KM, Ferrari BT. Treating the "Not-Invented-Here Syndrome" in medical leadership: learning from the insights of outside disciplines. Acad Med 2019 Oct;94(10):1416-1418. [doi: 10.1097/ACM.0000000000002860] [Medline: 31274525]
16. Rauwerdink A, Kasteleyn MJ, Chavannes NH, Schijven MP. The successes and lessons of a Dutch University Hospitals' eHealth program: an evaluation study protocol. Clinical eHealth 2021;4:30-36. [doi: 10.1016/j.ceh.2020.12.002]
17. Kristensen M, Høiberg L, Nøhr C. Updated mapping of telemedicine projects in Denmark. Stud Health Technol Inform 2019;257:223-228. [Medline: 30741200]
18. Eisenstein E, Juzwishin D, Kushniruk A, Nahm M. Defining a framework for health information technology evaluation. Stud Health Technol Inform 2011;164:94-99. [Medline: 21335694]
19. Brand SL, Quinn C, Pearson M, Lennox C, Owens C, Kirkpatrick T, et al. Building programme theory to develop more adaptable and scalable complex interventions: realist formative process evaluation prior to full trial. Evaluation 2018 Oct 08;25(2):149-170. [doi: 10.1177/1356389018802134]
20. Deliverables of the Citrien fund - Program eHealth. Citrien Fund - program eHealth. URL: https://www.citrienfonds-ehealth.nl/over/citrien-1-evidence-based-ehealth/proces-en-deliverables/ [accessed 2021-10-16]
21. Rauwerdink A, Kasteleyn MJ, Haafkens JA, Chavannes NH, Schijven MP, steering committee, of the Citrien fund program eHealth. A national eHealth vision developed by University Medical Centres: a concept mapping study. Int J Med Inform 2020 Jan;133:104032. [doi: 10.1016/j.ijmedinf.2019.104032] [Medline: 31778886]
22. Nepal S, Li J, Jang-Jaccard J, Alem L. A framework for telehealth program evaluation. Telemed J E Health 2014 Apr;20(4):393-404. [doi: 10.1089/tmj.2013.0093] [Medline: 24621384]
23. SurveyMonkey. URL: http://www.surveymonkey.com [accessed 2021-10-25]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. E-health toolkit. E-health Citrien programma. 2021. URL: https://www.citrienfonds-ehealth.nl/e-health-toolkit/
26. Koole MA, Kauw D, Winter MM, Dohmen DA, Tulevski II, de Haan R, et al. First real-world experience with mobile health telemonitoring in adult patients with congenital heart disease. Neth Heart J 2019 Jan;27(1):30-37 [FREE Full text] [doi: 10.1007/s12471-018-1201-6] [Medline: 30488380]
27. Kauw D, Koole MA, Winter MM, Dohmen DA, Tulevski II, Blok S, et al. Advantages of mobile health in the management of adult patients with congenital heart disease. Int J Med Inform 2019 Dec;132:104011 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104011] [Medline: 31654966]
28. Vreugdenhil MM, Ranke S, de Man Y, Haan MM, Kool RB. Patient and health care provider experiences with a recently introduced patient portal in an academic hospital in the Netherlands: mixed methods study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
29. Neijenhuijs KI, van der Hout A, Veldhuijzen E, Scholten-Peeters GG, van Uden-Kraan CF, Cuijpers P, et al. Translation of the eHealth impact questionnaire for a population of Dutch electronic health users: validation study. J Med Internet Res 2019 Aug 26;21(8):e13408 [FREE Full text] [doi: 10.2196/13408] [Medline: 31452516]
30. Breteler MJ, KleinJan E, Numan L, Ruurda JP, Van Hillegersberg R, Leenen LP, et al. Are current wireless monitoring systems capable of detecting adverse events in high-risk surgical patients? A descriptive study. Injury 2020 May;51 Suppl 2:S97-105 [FREE Full text] [doi: 10.1016/j.injury.2019.11.018] [Medline: 31761422]
31. Breteler M, KleinJan E, Dohmen D, Leenen LP, van Hillegersberg R, Ruurda JP, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology 2020 Mar;132(3):424-439 [FREE Full text] [doi: 10.1097/ALN.0000000000003029] [Medline: 31743149]
32. De Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020 Aug 14;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]
33. van den Heuvel JF, Lely AT, Huisman JJ, Trappenburg JC, Franx A, Bekker MN. SAFE@HOME: digital health platform facilitating a new care path for women at increased risk of preeclampsia - a case-control study. Pregnancy Hypertens 2020 Oct;22:30-36 [FREE Full text] [doi: 10.1016/j.preghy.2020.07.006] [Medline: 32717653]
34. Kubben P, Dumontier M, Dekker A. Fundamentals of Clinical Data Science. Cham: Springer International Publishing; 2018.
35. NHS digital. National Health Service. URL: https://digital.nhs.uk/ [accessed 2020-09-01]
36. Patient@Home. University of Southern Denmark. URL: https://www.en.patientathome.dk/about-patient@home.aspx [accessed 2020-09-01]
37. 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]
38. Liu C, Shao S, Liu C, Bennett GG, Prvu Bettger J, Yan LL. Academia-industry digital health collaborations: a cross-cultural analysis of barriers and facilitators. Digit Health 2019 Sep 26;5:2055207619878627 [FREE Full text] [doi: 10.1177/2055207619878627] [Medline: 31632684]

39. Vedlūga T, Mikulskienė B. Stakeholder driven indicators for eHealth performance management. Eval Program Plann 2017 Aug;63:82-92. [doi: 10.1016/j.evalprogplan.2017.03.001] [Medline: 28433891]
40. Citrien Fund 2019 - 2022: promoting sustainable healthcare. ZonMW. URL: https://www.zonmw.nl/nl/over-zonmw/e-health-en-ict-in-de-zorg/programmas/programma-detail/citrienfonds/ [accessed 2019-01-01]
41. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
42. Barsom EZ, Feenstra TM, Bemelman WA, Bonjer JH, Schijven MP. Coping with COVID-19: scaling up virtual care to standard practice. Nat Med 2020 May;26(5):632-634. [doi: 10.1038/s41591-020-0845-0] [Medline: 32405054]
43. The Lancet Digital Health. Pandemic versus pandemonium: fighting on two fronts. Lancet Digit Health 2020 Jun;2(6):e268 [FREE Full text] [doi: 10.1016/S2589-7500(20)30113-8] [Medline: 32501437]

Abbreviations

AI: artificial intelligence
CF: Citrien Fund
CSIRO: Commonwealth Scientific and Industrial Research Organization
DRIVE: digital, research, informatics, and virtual environments
IT: information technology
NHS: National Health Service

Edited by R Kukafka; submitted 28.10.20; peer-reviewed by E Metting, S Bond, S Pit; comments to author 05.12.20; revised version received 23.02.21; accepted 27.07.21; published 25.11.21.

Please cite as:
Rauwerdink A, Kasteleyn MJ, Chavannes NH, Schijven MP
Successes of and Lessons From the First Joint eHealth Program of the Dutch University Hospitals: Evaluation Study
J Med Internet Res 2021;23(11):e25170
URL: https://www.jmir.org/2021/11/e25170
doi:10.2196/25170
PMID:34842536

Original Paper

Information and Scientific Impact of Advanced Therapies in the Age of Mass Media: Altmetrics-Based Analysis of Tissue Engineering

Antonio Santisteban-Espejo[1,2], MD, PhD; Miguel Angel Martin-Piedra[3], DDS, PhD; Antonio Campos[3], MD, PhD; Julia Moran-Sanchez[4,5], MD; Manuel J Cobo[6], PhD; Ana I Pacheco-Serrano[7], PT; Jose A Moral-Munoz[2,7], PhD, PT

[1]Department of Pathology, Puerta del Mar University Hospital, Cadiz, Spain

[2]Institute of Research and Innovation in Biomedical Sciences of the Province of Cadiz (INiBICA), University of Cadiz, Cadiz, Spain

[3]Department of Histology, Tissue Engineering Group, University of Granada, Granada, Spain

[4]Department of Medicine, University of Cadiz, Cadiz, Spain

[5]Department of Hematology and Hemotherapy, Puerta del Mar University Hospital, Cadiz, Spain

[6]Department of Computer Science and Engineering, University of Cadiz, Cadiz, Spain

[7]Department of Nursing and Physiotherapy, University of Cadiz, Cadiz, Spain

Corresponding Author:
Miguel Angel Martin-Piedra, DDS, PhD
Department of Histology
Tissue Engineering Group
University of Granada
School of Medicine
Avda de la Ilustración, 11
Granada, 18016
Spain
Phone: 34 958241000 ext 40703
Email: mmartin@ugr.es

Abstract

Background: Tissue engineering (TE) constitutes a multidisciplinary field aiming to construct artificial tissues to regenerate end-stage organs. Its development has taken place since the last decade of the 20th century, entailing a clinical revolution. TE research groups have worked and shared relevant information in the mass media era. Thus, it would be interesting to study the online dimension of TE research and to compare it with traditional measures of scientific impact.

Objective: The objective of this study was to evaluate the online dimension of TE documents from 2012 to 2018 using metadata obtained from the Web of Science (WoS) and Altmetric and to develop a prediction equation for the impact of TE documents from altmetric scores.

Methods: We analyzed 10,112 TE documents through descriptive and statistical methods. First, the TE temporal evolution was exposed for WoS and 15 online platforms (news, blogs, policy, Twitter, patents, peer review, Weibo, Facebook, Wikipedia, Google, Reddit, F1000, Q&A, video, and Mendeley Readers). The 10 most cited TE original articles were ranked according to the normalized WoS citations and the normalized Altmetric Attention Score. Second, to better comprehend the TE online framework, correlation and factor analyses were performed based on the suitable results previously obtained for the Bartlett sphericity and Kaiser–Meyer–Olkin tests. Finally, the linear regression model was applied to elucidate the relation between academics and online media and to construct a prediction equation for TE from altmetrics data.

Results: TE dynamic shows an upward trend in WoS citations, Twitter, Mendeley Readers, and Altmetric Scores. However, WoS and Altmetric rankings for the most cited documents clearly differ. When compared, the best correlation results were obtained for Mendeley Readers and WoS (ρ=0.71). In addition, the factor analysis identified 6 factors that could explain the previously observed differences between academic institutions and the online platforms evaluated. At this point, the mathematical model constructed is able to predict and explain more than 40% of TE WoS citations from Altmetric scores.

Conclusions: Scientific information related to the construction of bioartificial tissues increasingly reaches society through different online media. Because the focus of TE research importantly differs when the academic institutions and online platforms

are compared, basic and clinical research groups, academic institutions, and health politicians should make a coordinated effort toward the design and implementation of adequate strategies for information diffusion and population health education.

(J Med Internet Res 2021;23(11):e25394) doi:10.2196/25394

KEYWORDS

advanced therapies; tissue engineering; scientometrics; altmetrics; online; web; communication of science

Introduction

Tissue engineering (TE) is a multidisciplinary field aiming to develop biological substitutes that can restore, maintain, or even improve the structure or functionality of damaged tissues [1]. Since its appearance in 1988 [2], TE has globally spread to improve current therapeutic approaches, entailing a revolution in health sciences [3]. In this sense, several TE devices have been employed in the treatment of damaged blood vessels [4], peripheral nerve injuries [5], chronic skin ulcerations [6], oral mucosal replacement [7,8] and corneal lesions [9].

The crescent interest and the fast development of TE have been demonstrated from a quantitative perspective by showing the incremental number of TE publications during the last decade [10]. Moreover, its cognitive and social frameworks have been described by means of science mapping analysis techniques [11]. These bibliometric-based studies can serve as a guide to help administrative authorities to better plan funding allocations and to promote synergies among research groups, as previously exhibited in other scientific areas [12,13].

In this sense, traditional bibliometric analysis employs the information extracted from academic documents (ie, citations or keywords) to comprehend the evolution of a scientific discipline, such as TE [10,11,14]. However, classical bibliometric methods have been largely reviewed because of their fewer adequacy to assess the real dimension of scientific enterprise and due to a relative inattention to the societal dimension of scientific endeavor [15]. Consequently, a new kind of metrics, called alternative metrics or altmetrics, has been proposed to obtain, evaluate, and characterize scientific information through data content in social media [16].

Altmetrics describes a web-based metrics used to understand the impact of publications and other scholarly materials by using data from social media platforms (ie, Twitter, Facebook, Google+, blogs, Mendeley Readers, CiteULike, Reddit, and Wikipedia, among others) [17]. The emergence and development of these metrics are related to the social media revolution: there are now different groups of the population, nonauthor professionals, which read research articles and also share them; furthermore, new types of academic outputs have appeared [18]. Hence, the traditional acceptance that scientific output is disseminated solely through academic media, such as journals, conferences, or specialized books, has now changed.

In addition, the online public nature of these metrics allows to track mentions of scholarly articles across the online landscape faster and broader than traditional citation metrics [19]. The validity and potential of altmetrics and its necessary collaborative relation with classical metrics have been demonstrated in several disciplines [20]. Motivations on the impact that these metrics could offer on professional research careers have been also scrutinized [21].

Then, within the context of a global science where information is shared and consumed in the web, even before its general validation for the scientific community, it would be interesting to explore the online dimension of a multidisciplinary and dynamic science such as TE. Among the recent advances in health sciences, the construction of biosimilar tissues constitutes one of the most powerful approaches to achieve the successful treatment of previously untreated conditions. To the best of our knowledge, there are no documents available that evaluate the online dimension of TE research since its appearance at the end of the 20th century. Thus, the primary aim of this study was to determine the characters of TE behavior online and to compare it with traditional metrics of scientific impact.

Methods

Sample

The metadata used in this study were obtained from the Web of Science (WoS) Core Collection bibliographic database. WoS is considered one of the most relevant scientific information sources, as it contains reliable evidence about citations, and is widely used in research evaluations [22].

The search strategy used in this study was “TISSUE ENGINEER*” or “TISSUE-ENGINEER*”, and it was applied on the Science Citation Index-Expanded Collection for a period between 2012 and 2018. We performed this search strategy to accurately discriminate between genuine TE documents and documents belonging to other related areas such as regenerative medicine or cellular therapy [23]. As originally described by Langer and Vacanti [1], TE is defined by the use of cell sources, matrices, and growing factors to construct biomimetic tissues with a therapeutic impact on human health [1], which differs from other emerging biomedical approaches based on the sole use of cultured stem cells or biomaterials without giving rise to a human bioartificial tissue. In this sense, our aim was to capture this precise notion of TE research.

Once the metadata were extracted, we excluded reviews, book chapters, meeting abstracts, and proceeding articles. Then, original articles obtained from this research were matched with the information available on Altmetric online [24], which holds important social information since 2012 from a much broader spectrum of sources than traditional metrics (eg, web-based references, news media mentions, Twitter mentions, or patents, among others) [25].

Descriptive Analysis

To comprehend the behavior of TE in the social web and to compare it with traditional metrics, we carried out 2 different

analyses. First, we evaluated the presence of original articles regarding TE in 7 different platforms (WoS, Altmetric Attention Score, Twitter, patents, Facebook, Mendeley Readers, and news) as the percentage of documents with at least one mention or a citation from 2012 to 2018. Following Eysenbach [26], in the case of Twitter, we called each mention a *tweetation*, which includes the mention of a TE journal article URL, retweet of the same tweet, or sending a modified tweet by other users [26]. In addition, we obtained the top 10 most cited TE original articles from 2012 to 2018 and ranked them according to 2 parameters: the normalized WoS citations and the normalized Altmetric Attention Score. Those measures were calculated using the rationale of the normalized citation impact. It was calculated by dividing the count of citing items by the average of citations for documents with the same year of publication in our corpus of documents. The Altmetric Attention Score has been previously employed as a bibliometric measure of online attention [25].

Statistical Analysis

To better characterize TE structure online, we performed 3 different statistical tests: Spearman correlation test [27], factor analysis [28], and linear regression model [29]. The collection of cites using traditional metrics requires several years, while the data provided by Altmetric before 2015 were not extensive, as the platform was only founded in 2012. For this reason, correlation and factor analyses were performed on publications retrieved from 2015 to 2018. This strategy has been used previously in other altmetrics studies [30]. Furthermore, all citation and mention counts were transformed with the formula Ln(1+x) before processing to reduce skewing [30].

To verify that the data set does not follow a normal distribution, the Kolmogorov–Smirnov test was performed for the next 16 variables that were evaluated, overall, to characterize the field: (1) WoS citations, (2) news, (3) blogs, (4) policy, (5) Twitter, (6) patents, (7) peer review, (8) Weibo, (9) Facebook, (10) Wikipedia, (11) Google, (12) Reddit, (13) F1000, (14) Q&A, (15) video, and (16) Mendeley Readers. The Spearman correlation was then obtained for the variables previously described, and the statistical significance was defined as $P<.05$.

Once the correlation data were obtained, factor analysis was performed. Factor analysis allowed us to identify the common variables or factors that could explain the previously observed correlation data. In this sense, Bartlett sphericity and Kaiser–Meyer–Olkin tests were performed prior to assessing the suitability of factor analysis [31]. Finally, the linear regression model was applied to obtain a mathematical expression of the influence of alternative metrics on a traditional measure of scientific impact such as the citation counts. The equation constructed contains a group of variables identified in the correlation and factor analyses, which allows us to predict the number of WoS citations in 2018 from 2015 TE Altmetric scores. Finally, the equation was used to calculate the predicted citation(s) of the documents published in 2015. A *t*-test analysis was employed to determine the significance (95% CI and significance at $P<.05$). JASP (freeware; University of Amsterdam, Amsterdam, The Netherlands) was employed to perform all the statistical analyses [32].

Results

Sample

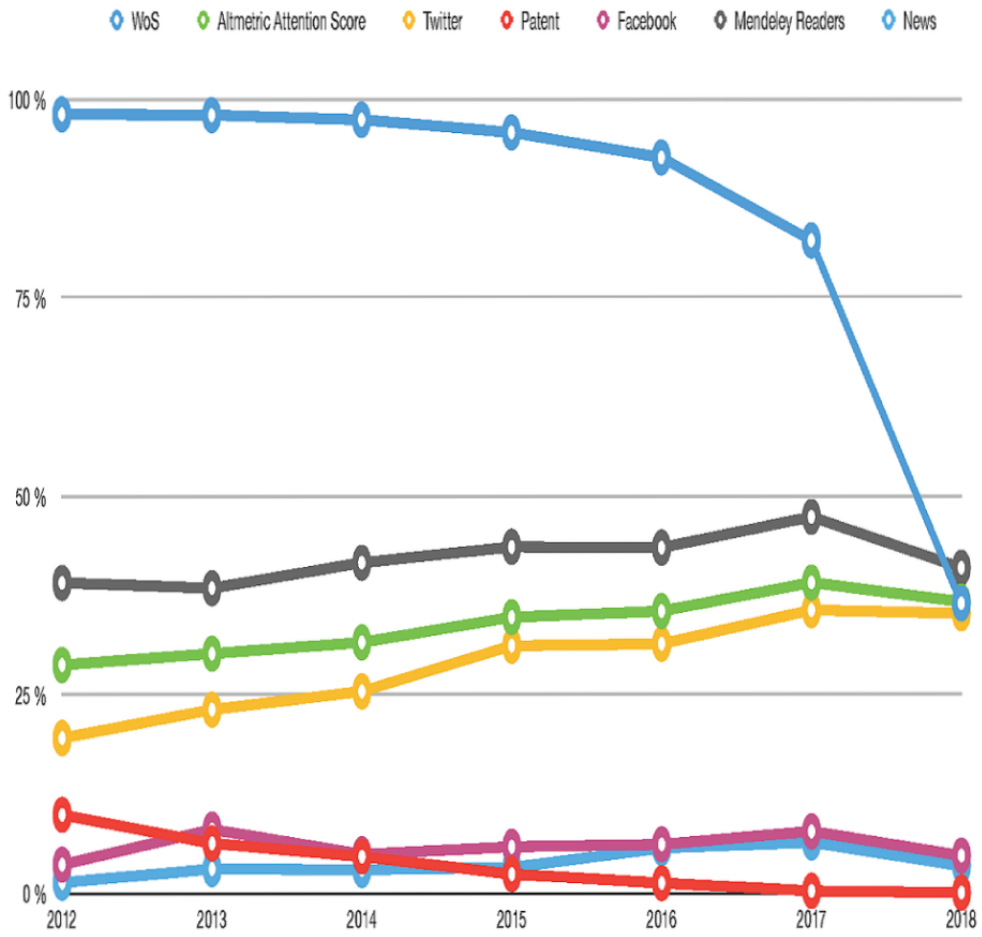
After performing the search strategy described, a total of 23,179 documents pertaining to TE were retrieved from WoS for the period from 2012 to 2018. A process of matching between the DOIs available in the WoS and the Altmetric data was then performed. Finally, a total of 10,112 documents (43.63%) with an Altmetric score of 1 or higher were obtained.

Descriptive Analysis

Evolution of TE Documents in WoS and the Online Web

The presence of TE documents in WoS and online is shown in Figure 1. The trend lines indicate the evolution of the percentage of documents with at least one citation or mention during the period 2012-2018. In WoS, the percentage of documents exceeds 85.00% from 2012 to 2017. However, the nearness of 2018 to the time of data acquisition explains the result of WoS citations in that year (40.93%) when these metadata were not already collected. The evolution of TE documents in Twitter, Mendeley Readers, and the Altmetric Attention Score shows an upward trend from the beginning of the period studied. In this sense, documents with at least one mention in the reference manager Mendeley Readers were close to the 50% in 2017. By contrast, the presence of TE documents in platforms such as Facebook, patents, and news was less than 10% for the whole period studied.

Figure 1. Percentage of documents with at least one citation/mention for the period 2012-2018. Only those platforms with more than 5% in any year were represented. WoS: Web of Science.



Ranking of TE Documents According to WoS Citations and Altmetric Attention Score

The top 10 TE documents ranked by their normalized WoS citations and normalized Altmetric Attention Score are presented in Tables 1 and 2.

Table 1. Top 10 tissue engineering documents ranked by WoS[a] citations for the period 2012-2018.

WoS rank	Altimetric rank	Normalized WoS citations	Normalized Altmetric Attention Score	Reference
1	5	38.32	140.76	[33]
2	21	35.19	54.11	[34]
3	43	22.19	33.31	[35]
4	602	21.49	2.16	[36]
5	291	19.78	5.73	[37]
6	6022	18.10	0.11	[38]
7	8060	18.10	0	[39]
8	2259	18.03	0.59	[40]
9	3738	16.97	0.22	[41]
10	8	14.91	68.29	[42]

[a]WoS: Web of Science.

Table 2. Top 10 tissue engineering documents ranked by Altmetric Attention Score for the period 2012-2018.

WoS[a] rank	Altimetric rank	Normalized WoS citations	Normalized Altmetric Attention Score	Reference
365	1	3.46	204.96	[43]
16	2	12.97	199.36	[44]
83	3	6.83	149.21	[45]
39	4	9.60	141.62	[46]
1	5	38.32	140.76	[33]
8853	6	0	115.27	[47]
1584	7	1.73	85.10	[48]
10	8	14.91	68.29	[49]
307	9	3.80	67.04	[50]
252	10	4.37	65.81	[51]

[a]WoS: Web of Science.

Tables 1 and 2 show a remarkable discrepancy between classical (normalized WoS citations) and alternative (normalized Altmetric Attention Score) metrics among the most valued documents.

On the one hand, the original article by Deng et al [39], reporting multifunctional stimuli-responsive hydrogels with self-healing, high conductivity, and rapid recovery through host–guest interactions, has a remarkable scholarly impact, being the 7th top-cited document when analyzing normalized WoS citations. However, the Altmetric Attention Score was null for this paper, suggesting that in vitro research could not attract as much societal attention as translational research. On the other hand, the research study by Nichols et al [47], regarding the transplantation of bioengineered lung into a large-animal model, employs a very translational approach to TE, and thus its social impact is reflected by the high Altmetric Attention Score, although its scholar relevance was not yet evident.

Statistical Analysis

Correlation Analysis

The results of the correlation analysis between traditional and alternative metrics of all retrieved publications from 2015 to 2018 are presented in Table 3.

Table 3. Spearman correlation results between pairs of variables for tissue engineering articles published from 2015 to 2018.[a]

	WoS[b]	News	Blogs	Policy	Twitter	Patents	Peer review	Weibo	Facebook	Wikipedia	Google	Reddit	F1000	Q&A	Video
WoS															
News	*0.144*														
Blogs	*0.137*	*0.387*													
Policy	0.049	0.065	0.064												
Twitter	*0.176*	*0.149*	*0.158*	0.009											
Patents	*0.114*	*0.093*	*0.09*	–0.009	–0.068										
Peer review	–0.006	–0.01	*0.136*	–0.001	–0.006	–0.009									
Weibo	0.04	*0.108*	*0.108*	–0.002	0.063	–0.011	–0.002								
Facebook	0.064	*0.2*	*0.213*	0.046	*0.213*	0.047	–0.014	*0.081*							
Wikipedia	0.076	0.072	0.073	–0.004	0.073	0.032	–0.004	*0.14*	0.039						
Google	0.071	*0.109*	*0.184*	*0.138*	0.073	0.013	–0.005	*0.11*	*0.184*	0.086					
Reddit	–0.012	0.008	0.028	–0.005	–0.015	–0.03	–0.005	*0.118*	0.016	–0.013	0.025				
F1000	0.054	0.036	0.063	–0.003	0.05	0.013	–0.003	*0.164*	0.015	0.063	0.046	–0.011			
Q&A	0.003	–0.007	–0.007	–0.001	–0.004	–0.006	–0.001	–0.001	–0.01	–0.003	–0.003	–0.003	–0.002		
Video	–0.033	0.008	0.036	–0.003	0.006	–0.02	–0.003	–0.004	0.061	–0.009	–0.011	–0.011	–0.008	–0.002	
Mendeley Readers	*0.716*	*0.198*	*0.197*	0.021	*0.243*	*0.104*	–0.025	0.049	*0.107*	0.077	*0.098*	0.012	0.067	0.008	0.03

[a]Italicized values mean $P<.05$.

[b]WoS: Web of Science.

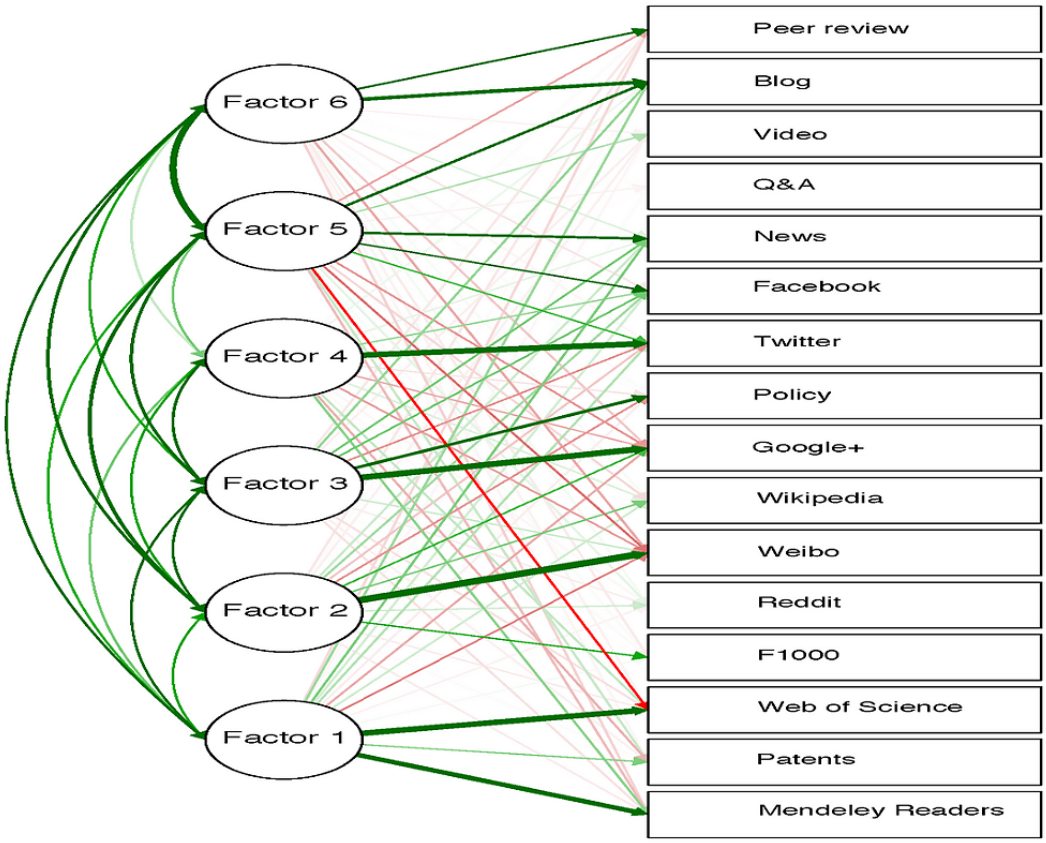
Overall, the number of citations on Mendeley Readers and WoS shows the best correlation results (ρ=0.71) and platforms such as Twitter (ρ=0.17) and news (ρ=0.14) have a suitable correlation. However, the correlation results obtained for Wikipedia, Facebook, F1000 mentions, and Q&A mentions were weak, and an inverse correlation was observed for TE documents appearing in 3 online platforms: peer review mentions (ρ=−0.006), Reddit mentions (ρ=−0.01), and video mentions (ρ=−0.03).

Factor Analysis

First, a value of 5629.85 ($P<.001$) for a chi-square approximation of the Bartlett sphericity test and a value of 0.700 for the Kaiser–Meyer–Olkin test confirmed the suitability of factor analysis. Then, factor analysis identified 6 different components or factors that could explain the correlation results. These factors are shown in Figure 2 and labeled as F1, F2, F3, F4, F5, and F6. Positive and negative results are indicated with green and red lines, respectively.

F1 exposes the relation between WoS citations and Mendeley Readers. However, the remaining factors (F2-F6) most likely account for a different type of scientific impact not directly associated with TE professional researchers and readers. Regarding this, F3 acts as a common factor for Google and policy mentions, and, interestingly, 3 social platforms (blogs, news, and Facebook) appear together within F5. Finally, the mentions of TE documents in Twitter (*tweetations*) were strongly tied with a unique factor (F4), suggesting a particular behavior and structure for TE information shared in Twitter.

Figure 2. Results of the factor analysis for the 2015 production in the research field of tissue engineering.



Linear Regression Analysis

The correlation coefficient (r) and the determination coefficient (R^2) obtained were equal to 0.645 and 0.414, respectively. In addition, the statistical test for the analysis of variance was significant (P<.001). Consequently, the mathematical model constructed explains more than the 40% of the variation in the number of WoS citations obtained for TE documents in 2018 from 2015 Altmetric scores.

The variable Mendeley Readers constitutes the best citation predictor for TE documents as it holds the higher result for r (r=0.599). The rest of the Altmetric scores also had a positive correlation but the strength of the observed association was weaker.

The prediction equation for 2018 TE WoS citation counts from 2015 Altmetric scores can be expressed as follows:

Ln (1 + WoS)=–27.25 + 5.37 x Ln (1 + blog) + 0.82 × Ln (1 + news) + 12.78 × Ln (1 + Mendeley Readers) + 5.83 × Ln (1 + patent) + 0.75 × Ln (1 + Twitter).

Finally, no significant differences were found (P=.12) for the predicted citations versus the real citations rates for the documents published in 2015.

Discussion

Principal Findings

The seminal article published by Langer and Vacanti [1] laid the foundations for TE. Since then, TE has evolved and given rise to an interdisciplinary field that applies the principles of engineering and life sciences toward the development of biological substitutes that can restore, maintain, or even improve tissue functions. Within contemporary medicine, TE is considered one of the most promising advanced therapies, as it has the potential to overcome traditional problems associated with organ failure and to treat previously untreated conditions [52]. Its onset and application to the clinical practice have led to a revolution in surgery and transplantation procedures, as new bioartificial tissue devices are now available for therapy with a considerably less risk of infection transmission and immune-mediated organ rejection [53].

As a consequence, a crescent interest has appeared, aiming to elucidate global trends in TE and its cognitive and social framework [10,11]. These bibliometrics-based approaches utilize, in common, the traditional measures of scientific impact, such as citations and publications. Moreover, the social maps and conceptual diagrams proposed suffer from the same bias, as both the relations among institutions and the key notions identified are based on the number of citations and co-occurrence of keywords [54]. Thus, new approaches are needed to better characterize and comprehend the real impact of TE in our society. In this sense, the association of traditional

bibliometrics with alternative metrics (altmetrics) could render a more sensible and realistic view of TE behavior nowadays [55].

Hence, in this study, we have carried out an altmetrics-based analysis of the core documents of TE retrieved from WoS between 2012 and 2018. We have previously employed this query term to analyze the global trends of TE [10], the cognitive and social framework of TE [11], and the structure and evolution of TE reviews [56], in an attempt to replicate the same search strategy highlighting the value of the reproducibility and comparability of our results. To our best knowledge, there is no previous literature that defines TE structure and its major characters online or its essential divergence with other widespread platforms in clinical medicine such as scientific journals.

In this regard, we first performed a descriptive analysis of evolution of TE documents in WoS and 6 different web-based platforms (ie, Facebook, patents, Twitter, news, Mendeley Readers, and Altmetric Attention Score). The presence of TE documents in WoS is significant over the rest, suggesting the existence of a well-established research dynamic where academic and professional health practitioners collect and consult applicable clinical information in renowned databases. Besides, TE diffusion in Twitter stood out within the group of social networks consulted; it is interesting to note a growing trend for the whole period evaluated, and a particular pattern of scientific information diffusion in Twitter could explain these results.

On the one hand, the own structure of Twitter, a micro-blogging platform that enables the users to “tweet” short messages with their virtual colleagues, has developed a singular model of scientific communication and a special information flow [57,58]. Kwak et al [58] demonstrated that retweets constitute the nucleus of this original model. Hence, retweets of TE documents could spread their information beyond the limits of their original authors, expanding them to the broad space of the followers’ networks [59,60]. In addition, relevant information about new TE devices may reach primary care physicians and groups of patients through this network, optimizing the communication between different health care levels and the education of society [61]. Eysenbach [26] reported that highly tweeted articles are 11 times more likely to end up as being highly cited and that Tweets correlate with traditional metrics of scientific impact [26]. Consequently, the upward trend of TE documents in Twitter could also be explained in terms of this higher academic impact.

To better comprehend the similarities and differences between the focus of TE documents online and in traditional scholar media, we identified the 10 most cited TE documents from 2012 to 2018. We then ranked and compared them according to the number of normalized WoS citations and the normalized Altmetric Attention Score. The results obtained demonstrated a clear discrepancy between the rankings of TE documents, suggesting that citations in WoS and interests of online users do not follow the same path. Differences between metrics tend to be more remarkable when comparing the top-ranked documents for each metric. This comparison, although cannot be used for validation purposes, is useful to elucidate this differential pattern. This kind of dissimilar relation, where scholar- and web-based attention clearly differs, has never been demonstrated for TE as a discipline, although it is not exclusive of it. In this sense, similar results have been shown in other research fields, revealing that social and academic assumptions of scientific advances are not guided by identical principia [62,63].

This finding is not a negative result but rather a consequence of the varying nature of traditional and alternative metrics, as well as the social and dynamic context in which research takes place. It has been reported that, in medical and applied sciences, an important share of information targets is found outside the research and scholar community and that traditional citations are only partial measures of impact and use of information [64]. In accordance with Bornmann [65], citations only assess the impact of scholarly literature on those who cite, and this neglects many audiences of scholarly literature who may read the paper, but do not cite it as “pure” readers [65]. Furthermore, the task of assessing the impact of science has to take into account some policy and society demands. These societal, policy-driven, and technical demands have led to the emergence of altmetrics as an evolved methodology to broaden the impact of research on both researchers and policy demands as promoters of research and society as final users of developed technology through advances in research.

As TE is devoted to the construction of biomimetic tissues that can restore, maintain, or even improve the structure or functionality of damaged tissues [3], and to treat previously untreated conditions [66,67], its social demands are particularly important [68]. In this sense, the use of altmetrics, combined with classical measures of scientific impact, could provide a wider context on the real influence of TE research in society.

In addition to the descriptive analysis, we applied 3 different statistical tests: Spearman correlation, factor analysis, and the linear regression model. The correlation study showed that TE citations in WoS and the number of readers in Mendeley Readers have the highest value (ρ=0.71). This finding can be explained by attending to the own nature of Mendeley Readers, as it is a citation manager tool essentially used to store and share references by a community of bibliographic users. The use of Mendeley Readers has been previously correlated with future citation counts in several biomedical sciences fields [69]. In this way, citations of TE documents in WoS are equally well-correlated with the number of Mendeley Readers. Because TE researchers could use the previously stored documents as cited documents for their own future publications, the correlation results are, to some extent, explainable. However, Mendeley Readers users do not have to be publishing academics exclusively, and may also be practitioners or students, as previously demonstrated [70,71]. Therefore, the correlation observed in TE research should be related to a broader spectrum of scientific activity and not just restricted to experts and research groups that publish in specialized journals.

A positive but weaker correlation was obtained for online platforms such as Twitter, news, and blogs. However, the mention of TE documents in video and Reddit is lesser, because

of an inverse correlation. These results are most likely influenced by the structure and the type of readers on these platforms. For example, in Reddit, virality constitutes a crucial factor [72]. As stated by Berger and Milkman [73], those contents that evoke emotions of activation (eg, anger, awe, anxiety) are more suitable to become viral, in contrast to deactivating emotions (eg, softness) [73]. Hence, documents referring to the construction of bioartificial tissues could be mentioned in Reddit to be criticized or report findings that are surprising and shocking for common readers, but not so relevant for a specialized audience.

For instance, correlation studies could obscure the genuine relationships existing between a set of variables. This potential bias is particularly important when a predominant or strong association exists [30]. In this sense, factor analysis could serve to identify the common factors or components that explain the previously observed correlation. In this regard, factor analysis of TE production showed the existence of 6 differentiated factors (F1-F6).

F1 is tied to readers in Mendeley Readers, citations in WoS, and patents. Interestingly, the final goal that guides TE research is the clinical application of bioengineered tissue devices in the daily practice of the medical specialties. For this achievement, 2 previous requirements must be guaranteed: the communication of the scientific results in a peer-reviewed journal and the acquisition of a patent license. As this process is causally related to the employment of citation manager and paper collection in well-known databases, factor analysis reveals consistent results. F2 (Weibo, F100, Reddit, and Wikipedia) probably accounts for a different kind of TE information consumption. A more informal communication of results with less scientific rigor mostly presided over the components that integrate this factor.

The relation between policy and Goggle in F3 is not clear, as the latter can be used to filter and obtain a heterogeneous and vast amount of information related to TE, and not just the legal requirements for TE application in clinics. It is interesting to note that Twitter acquires an individual dimension in F4, constituting a social network distinguished from the rest. Nevertheless, news and Facebook appear together in F5 and blogs constitute a component of F6. A plausible explanation for this leading role of Twitter in the diffusion of TE information is that the development of TE has taken place in parallel with the burst of social media. Probably, as previously stated for other scientific disciplines, TE researchers have substituted the idea of academic community for the virtual department [74,75]. Moreover, the structural multidisciplinary nature of TE and the relationships between the biomaterials industry, research groups, and clinicians can be ideally displayed using a social network such as Twitter [10,58,76].

Finally, we aimed to develop a mathematical model for TE documents to predict the influence of Altmetric scores on future citation counts with relative accuracy. However, in accordance with Thelwall and Nevill [30], it is reasonable to consider alternative metrics in conjunction with journal impact to get an idea about which articles are more likely to attract longer-term citations [30]. Applying this logic to TE production, we established a linear regression equation to derive 2018 TE citation counts from 2015 Altmetric indexes. The model is able to explain more than the 40% of variation in the number of WoS citations for TE documents that Altmetric tracked (R^2=49.6%); regression results were statistically significant, and so the association between measures such as publications or citations and the impact of scientific work online could serve to better characterize the movement of information in biomedical disciplines, such as TE.

We hope this article serves to stimulate the adequate use of web-based platforms in the communication and diffusion of scientific information in TE. We are firmly convinced that, as wisely stated by Weigold [77], the sharing of well-constructed information online contributes to informing society about real possibilities of scientific progress.

Limitations

Although the findings provided in this study are interesting, several limitations must be addressed. First, only a percentage of the publications indexed in WoS are available on Altmetric, and consequently, the conclusions of the study are influenced by the core of documents obtained. Second, the factor analysis is performed for only 1 year; although the behavior of the research area could be similar, it could be influenced by the published topics or other factors. Finally, the intentional tweeting by the publisher or the editor of the journal was not analyzed.

Furthermore, the use of altmetrics lead to some potential disadvantages, especially when they are used as the only indicator for impact assessment. There is also the difficulty with field normalization, which makes it difficult to compare the impact of different disciplines [78]. Besides, altmetrics could be affected by an incomplete and biased coverage of impact areas (eg, most Chinese regions do not use Twitter), which makes it difficult to compare the impact of different regions [79]. Importantly, altmetrics present a lack of quality control, and as such they are susceptible to deliberate or accidental manipulation, which may promote sensational outcomes, and the subsequent loss of credibility, if they are used as a sole indicator for impact assessment [80]. However, some of these drawbacks can be controlled when analyzing a large set of documents, and in this sense alternative metrics seem to be more prevalent and useful in health sciences when compared with other fields [81,82]. Thus, we consider that new alternative metrics are not replacing the classical ones. Indeed, these are 2 different approaches with a common goal: traditional metrics attempt to assess the scholarly impact among researchers, whereas alternative metrics try to evaluate policy and societal demands on a specific scientific issue. Although these approaches are different, they are positively correlated [26]. A randomized controlled trial [83] reported a causal relationship between the dissemination of research results through a web-based platform and subsequent citations.

Another limitation of this study is that we have restricted our search strategy to WoS, without exploring the presence of TE in other databases such as Scopus or Medline or employing a broader search strategy as reported in other studies [84]. However, WoS covers more than 250 scientific disciplines and

its total number of records is over 90 million [85]. When performing bibliometric analysis, citation data provided by WoS are considered one of its main advantages in comparison with other databases. Furthermore, the coverage of TE documents is not limited by the date of WoS construction (1960s), as the seminal paper on TE was published in 1993 [1].

Comparison With Prior Work

Previous studies of our group have described the global trends [10] and identified the cognitive and social framework of TE [11]. However, to our knowledge, this is the first study analyzing the online social dimension of TE as a research field in the age of mass media.

Summary of Findings

1. Online social media play a key role in the dissemination of information about advanced therapies and TE from academics to patients and health consumers.
2. The focus of TE research groups at the academic level and the most shared articles in the online mass media are not the same, as the ranking of the top 10 most cited TE documents in terms of normalized WoS citations and normalized Altmetric Attention Score were not homogeneous.
3. Mathematical models established based on information retrieved from alternative metrics (altmetrics) can be used to predict the impact of TE documents on citation counts.
4. Different actors (academics, groups of basic and translational researchers, health clinicians, data managers, and health information workers) should implement knowledge diffusion models about advanced therapies and TE in the online mass media.

Conclusions

TE has supposed a revolution in daily medical practice as tissue constructs are now available to treat severe conditions that previously remained untreated. Therefore, these new medical approaches have an impact on the population that can now be measured by altmetrics. These metrics differ from the classical academic metrics, but the knowledge of their influence on the final citation count could form the basis of different institutional or personal decision processes. The different actors involved in the scientific diffusion of the TE can use the results of this study to increase their interest in the use of social media and other online platforms as a window to the world, with the intention of reaching not only the scientific community, but also the general society.

Acknowledgments

This study was supported by CTS-115 (Tissue Engineering Research Group, University of Granada) from Junta de Andalucia, Spain, the Spanish State Research Agency through the project PID2019-105381GA-I00/AEI/10.13039/501100011033 (iScience), a postdoctoral grant (RH-0145-2020) from the Andalusia Health System, and the European Union Fondo Europeo de Desarrollo Regional para la Inversion Territorial Integrada Grant for Cadiz Province (PI-0032-2017). The authors thank Altmetric LLP (London, UK) for granting access to Altmetric Explorer for research purposes.

Authors' Contributions

MAM-P and JAM-M were responsible for conceptualization of this work. AS-E and JAM-M took care of the implemented methodology. MJC performed software analysis. AS-E, MAM-P, and JAM-M. validated the content. AS-E managed the formal analysis, performed the investigation, and managed resources. JM-S, MJC, and AIP-S performed data curation. JM-S and AIP-S were responsible for writing—original draft, while AS-E, MAM-P, and JAM-M were responsible for writing—review and editing. JAM-M took care of visualization. MAM-P, JAM-M, and AC supervised the work. AC was responsible for project administration. AC and MJC handled funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Langer R, Vacanti JP. Tissue engineering. Science 1993 May 14;260(5110):920-926. [doi: 10.1126/science.8493529] [Medline: 8493529]
2. Viola J, Bal B, Grad O. The Emergence of Tissue Engineering as a Research Field. URL: https://www.nsf.gov/pubs/2004/nsf0450/start.htm [accessed 2021-11-02]
3. Kaul H, Ventikos Y. On the genealogy of tissue engineering and regenerative medicine. Tissue Eng Part B Rev 2015 Apr;21(2):203-217 [FREE Full text] [doi: 10.1089/ten.TEB.2014.0285] [Medline: 25343302]
4. Kumar VA, Brewster LP, Caves JM, Chaikof EL. Tissue Engineering of Blood Vessels: Functional Requirements, Progress, and Future Challenges. Cardiovasc Eng Technol 2011 Sep 01;2(3):137-148 [FREE Full text] [doi: 10.1007/s13239-011-0049-3] [Medline: 23181145]
5. Carriel V, Alaminos M, Garzón I, Campos A, Cornelissen M. Tissue engineering of the peripheral nervous system. Expert Rev Neurother 2014 Mar;14(3):301-318. [doi: 10.1586/14737175.2014.887444] [Medline: 24506662]

6. Debels H, Hamdi M, Abberton K, Morrison W. Dermal matrices and bioengineered skin substitutes: a critical review of current options. Plast Reconstr Surg Glob Open 2015 Jan;3(1):e284 [FREE Full text] [doi: 10.1097/GOX.0000000000000219] [Medline: 25674365]
7. Sanchez-Quevedo MC, Alaminos M, Capitan LM, Moreu G, Garzon I, Crespo PV, et al. Histological and histochemical evaluation of human oral mucosa constructs developed by tissue engineering. Histol Histopathol 2007 Jun;22(6):631-640. [doi: 10.14670/HH-22.631] [Medline: 17357093]
8. Martín-Piedra MA, Alaminos M, Fernández-Valadés-Gámez R, España-López A, Liceras-Liceras E, Sánchez-Montesinos I, et al. Development of a multilayered palate substitute in rabbits: a histochemical ex vivo and in vivo analysis. Histochem Cell Biol 2017 Mar;147(3):377-388. [doi: 10.1007/s00418-016-1489-5] [Medline: 27600719]
9. Rico-Sánchez L, Garzón I, González-Andrades M, Ruíz-García A, Punzano M, Lizana-Moreno A, et al. Successful development and clinical translation of a novel anterior lamellar artificial cornea. J Tissue Eng Regen Med 2019 Dec;13(12):2142-2154 [FREE Full text] [doi: 10.1002/term.2951] [Medline: 31373143]
10. Santisteban-Espejo A, Campos F, Martin-Piedra L, Durand-Herrera D, Moral-Munoz JA, Campos A, et al. Global Tissue Engineering Trends: A Scientometric and Evolutive Study. Tissue Eng Part A 2018 Oct;24(19-20):1504-1517. [doi: 10.1089/ten.TEA.2018.0007] [Medline: 29687749]
11. Santisteban-Espejo A, Campos F, Chato-Astrain J, Durand-Herrera D, García-García O, Campos A, et al. Identification of Cognitive and Social Framework of Tissue Engineering by Science Mapping Analysis. Tissue Eng Part C Methods 2019 Jan;25(1):37-48. [doi: 10.1089/ten.TEC.2018.0213] [Medline: 30526420]
12. Abramo G, D'Angelo CA, Caprasecca A. Allocative efficiency in public research funding: Can bibliometrics help? Research Policy 2009 Feb;38(1):206-215. [doi: 10.1016/j.respol.2008.11.001]
13. Gheisari Y, Baharvand H, Nayernia K, Vasei M. Stem cell and tissue engineering research in the Islamic republic of Iran. Stem Cell Rev Rep 2012 Sep;8(3):629-639. [doi: 10.1007/s12015-011-9343-6] [Medline: 22350456]
14. Dai G, Yang KH, Li QH. [Bibliometric analysis on tissue engineering research literatures]. Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi 2000 Sep;14(5):308-310. [Medline: 12516484]
15. Powell AGMT, Bevan V, Brown C, Lewis WG. Altmetric Versus Bibliometric Perspective Regarding Publication Impact and Force. World J Surg 2018 Sep;42(9):2745-2756 [FREE Full text] [doi: 10.1007/s00268-018-4579-9] [Medline: 29536144]
16. Priem J, Piwowar H, Hemminger B. Altmetrics in the wild: an exploratory study of impact based on social media. arXiv Preprint posted online on March 20, 2012. [FREE Full text]
17. Veeranjaneyulu K. Altmetrics: new tools to measure research impact in the digitally networked environment. 2018 Jun 21 Presented at: National Conference of Agricultural Librarians and User Community (NCALUC-2017); 2018; Hisar, Haryana p. 495-505 URL: https://www.researchgate.net/publication/325619882_ALTMETRICS_NEW_TOOLS_TO_MEASURE_RESEARCH_IMPACT_IN_THE_DIGITALLY_NETWORKED_ENVIRONMENT
18. Moral-Munoz J, Cobo M. Measuring the online attention of the Rehabilitation Web of Science category : an Altmetrics-based analysis. 2018 Sep 14 Presented at: 23rd International Conference on Science and Technology Indicators; September 12-14, 2018; Leiden, The Netherlands p. 2018 URL: https://www.researchgate.net/publication/327594583_Measuring_the_online_attention_of_the_Rehabilitation_Web_of_Science_category_an_Altmetrics-based_analysis
19. Verma S, Margam M, Abu K. Altmetrics: A new way of assessing research impact beyond citations. 2018 Presented at: Changing Digital Landscape in Smart Environment; February 11, 2018; Delhi, India URL: https://www.researchgate.net/publication/325465926_Altmetrics_A_new_way_of_assessing_research_impact_beyond_citations
20. Haustein S, Bowman T, Costas R. Interpreting 'altmetrics': viewing acts on social media through the lens of citation and social theories. In: Theories of Informetrics and Scholarly Communication. Munich, Germany: De Gruyter Saur; 2016:372.
21. Jacob KJ, Longstaff H, Scott CT, Illes J. Focus on People and the Science Will Follow: Motivating Forces for Professional Movement in Stem Cell Research. Stem Cell Rev Rep 2015 Aug;11(4):533-539. [doi: 10.1007/s12015-015-9595-7] [Medline: 26022505]
22. Bakkalbasi N, Bauer K, Glover J, Wang L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. Biomed Digit Libr 2006 Jun 29;3(1):7 [FREE Full text] [doi: 10.1186/1742-5581-3-7] [Medline: 16805916]
23. Katari R, Peloso A, Orlando G. Tissue engineering and regenerative medicine: semantic considerations for an evolving paradigm. Front Bioeng Biotechnol 2014;2:57 [FREE Full text] [doi: 10.3389/fbioe.2014.00057] [Medline: 25629029]
24. Altmetric. URL: https://www.altmetric.com/ [accessed 2021-11-02]
25. Trueger NS, Thoma B, Hsu CH, Sullivan D, Peters L, Lin M. The Altmetric Score: A New Measure for Article-Level Dissemination and Impact. Ann Emerg Med 2015 Nov;66(5):549-553. [doi: 10.1016/j.annemergmed.2015.04.022] [Medline: 26004769]
26. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
27. Asuero AG, Sayago A, González AG. The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry 2007 Jan 12;36(1):41-59. [doi: 10.1080/10408340500526766]
28. Taherdoost H, Sahibuddin S, Jalaliyoon N. Exploratory factor analysis: concepts and theory. In: Advances in Applied and Pure Mathematics. Athens, Greece: WSEAS Press; 2014:375-382.

29. Thelwall M, Wilson P. Regression for citation data: An evaluation of different methods. Journal of Informetrics 2014 Oct;8(4):963-971. [doi: 10.1016/j.joi.2014.09.011]
30. Thelwall M, Nevill T. Could scientists use Altmetric.com scores to predict longer term citation counts? Journal of Informetrics 2018 Feb;12(1):237-248. [doi: 10.1016/j.joi.2018.01.008]
31. Chan LL, Idris N. Validity and Reliability of The Instrument Using Exploratory Factor Analysis and Cronbach's alpha. IJARBSS 2017 Oct 31;7(10):400-410. [doi: 10.6007/ijarbss/v7-i10/3387]
32. Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen J, et al. : Graphical Statistical Software for Common Statistical Designs. J. Stat. Soft 2019;88(2):1-17. [doi: 10.18637/jss.v088.i02]
33. Kang HW, Lee SJ, Ko IK, Kengla C, Yoo JJ, Atala A. A 3D bioprinting system to produce human-scale tissue constructs with structural integrity. Nat Biotechnol 2016 Mar;34(3):312-319. [doi: 10.1038/nbt.3413] [Medline: 26878319]
34. Sun JY, Zhao X, Illeperuma WRK, Chaudhuri O, Oh KH, Mooney DJ, et al. Highly stretchable and tough hydrogels. Nature 2012 Sep 06;489(7414):133-136 [FREE Full text] [doi: 10.1038/nature11409] [Medline: 22955625]
35. Kolesky DB, Truby RL, Gladman AS, Busbee TA, Homan KA, Lewis JA. 3D bioprinting of vascularized, heterogeneous cell-laden tissue constructs. Adv Mater 2014 May 21;26(19):3124-3130. [doi: 10.1002/adma.201305506] [Medline: 24550124]
36. Menasché P, Vanneaux V, Hagège A, Bel A, Cholley B, Parouchev A, et al. Transplantation of Human Embryonic Stem Cell-Derived Cardiovascular Progenitors for Severe Ischemic Left Ventricular Dysfunction. J Am Coll Cardiol 2018 Jan 30;71(4):429-438 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.047] [Medline: 29389360]
37. Gross BC, Erkal JL, Lockwood SY, Chen C, Spence DM. Evaluation of 3D printing and its potential impact on biotechnology and the chemical sciences. Anal Chem 2014 Apr 01;86(7):3240-3253. [doi: 10.1021/ac403397r] [Medline: 24432804]
38. Lai Y, Cao H, Wang X, Chen S, Zhang M, Wang N, et al. Porous composite scaffold incorporating osteogenic phytomolecule icariin for promoting skeletal regeneration in challenging osteonecrotic bone in rabbits. Biomaterials 2018 Jan;153:1-13. [doi: 10.1016/j.biomaterials.2017.10.025] [Medline: 29096397]
39. Deng Z, Guo Y, Zhao X, Ma PX, Guo B. Multifunctional Stimuli-Responsive Hydrogels with Self-Healing, High Conductivity, and Rapid Recovery through Host–Guest Interactions. Chem. Mater 2018 Feb 10;30(5):1729-1742. [doi: 10.1021/acs.chemmater.8b00008]
40. Croisier F, Jérôme C. Chitosan-based biomaterials for tissue engineering. European Polymer Journal 2013 Apr 10;49(4):780-792. [doi: 10.1016/j.eurpolymj.2012.12.009]
41. Ribeiro C, Costa CM, Correia DM, Nunes-Pereira J, Oliveira J, Martins P, et al. Electroactive poly(vinylidene fluoride)-based structures for advanced applications. Nat Protoc 2018 Apr;13(4):681-704. [doi: 10.1038/nprot.2017.157] [Medline: 29543796]
42. Workman MJ, Mahe MM, Trisno S, Poling HM, Watson CL, Sundaram N, et al. Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system. Nat Med 2017 Jan;23(1):49-59 [FREE Full text] [doi: 10.1038/nm.4233] [Medline: 27869805]
43. Gershlak JR, Hernandez S, Fontana G, Perreault LR, Hansen KJ, Larson SA, et al. Crossing kingdoms: Using decellularized plants as perfusable tissue engineering scaffolds. Biomaterials 2017 May;125:13-22 [FREE Full text] [doi: 10.1016/j.biomaterials.2017.02.011] [Medline: 28222326]
44. Laronda MM, Rutz AL, Xiao S, Whelan KA, Duncan FE, Roth EW, et al. A bioprosthetic ovary created using 3D printed microporous scaffolds restores ovarian function in sterilized mice. Nat Commun 2017 May 16;8:15261 [FREE Full text] [doi: 10.1038/ncomms15261] [Medline: 28509899]
45. Nawroth JC, Lee H, Feinberg AW, Ripplinger CM, McCain ML, Grosberg A, et al. A tissue-engineered jellyfish with biomimetic propulsion. Nat Biotechnol 2012 Aug;30(8):792-797 [FREE Full text] [doi: 10.1038/nbt.2269] [Medline: 22820316]
46. Park SJ, Gazzola M, Park KS, Park S, Di Santo V, Blevins EL, et al. Phototactic guidance of a tissue-engineered soft-robotic ray. Science 2016 Jul 08;353(6295):158-162 [FREE Full text] [doi: 10.1126/science.aaf4292] [Medline: 27387948]
47. Nichols JE, La Francesca S, Niles JA, Vega SP, Argueta LB, Frank L, et al. Production and transplantation of bioengineered lung into a large-animal model. Sci Transl Med 2018 Aug 01;10(452):eaao3926. [doi: 10.1126/scitranslmed.aao3926] [Medline: 30068570]
48. Moutos FT, Glass KA, Compton SA, Ross AK, Gersbach CA, Guilak F, et al. Anatomically shaped tissue-engineered cartilage with tunable and inducible anticytokine delivery for biological joint resurfacing. Proc Natl Acad Sci U S A 2016 Aug 02;113(31):E4513-E4522 [FREE Full text] [doi: 10.1073/pnas.1601639113] [Medline: 27432980]
49. Workman MJ, Mahe MM, Trisno S, Poling HM, Watson CL, Sundaram N, et al. Engineered human pluripotent-stem-cell-derived intestinal tissues with a functional enteric nervous system. Nat Med 2017 Jan;23(1):49-59 [FREE Full text] [doi: 10.1038/nm.4233] [Medline: 27869805]
50. Johnson BN, Lancaster KZ, Zhen G, He J, Gupta MK, Kong YL, et al. 3D Printed Anatomical Nerve Regeneration Pathways. Adv Funct Mater 2015 Oct 21;25(39):6205-6217 [FREE Full text] [doi: 10.1002/adfm.201501760] [Medline: 26924958]
51. Lu TY, Lin B, Kim J, Sullivan M, Tobita K, Salama G, et al. Repopulation of decellularized mouse heart with human induced pluripotent stem cell-derived cardiovascular progenitor cells. Nat Commun 2013;4:2307 [FREE Full text] [doi: 10.1038/ncomms3307] [Medline: 23942048]

52. Vacanti CA. The history of tissue engineering. J Cell Mol Med 2006;10(3):569-576 [FREE Full text] [doi: 10.1111/j.1582-4934.2006.tb00421.x] [Medline: 16989721]
53. Crupi A, Costa A, Tarnok A, Melzer S, Teodori L. Inflammation in tissue engineering: The Janus between engraftment and rejection. Eur J Immunol 2015 Dec;45(12):3222-3236 [FREE Full text] [doi: 10.1002/eji.201545818] [Medline: 26558332]
54. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
55. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. BMJ 2000 Apr 22;320(7242):1107-1111 [FREE Full text] [doi: 10.1136/bmj.320.7242.1107] [Medline: 10775218]
56. Martin-Piedra MA, Santisteban-Espejo A, Moral-Munoz JA, Campos F, Chato-Astrain J, Garcia-Garcia OD, et al. An Evolutive and Scientometric Research on Tissue Engineering Reviews. Tissue Eng Part A 2020 May;26(9-10):569-577. [doi: 10.1089/ten.TEA.2019.0247] [Medline: 31724488]
57. Finin T, Tseng B, Akshay J, Xiaodan S. Why We Twitter: Understanding Microblogging Usage and Communities. 2007. URL: https://ebiquity.umbc.edu/_file_directory_/papers/369.pdf [accessed 2021-11-02]
58. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web. 2010 Apr 23 Presented at: 19th International Conference on World Wide Web; April 26-30, 2010; Raleigh, NC. [doi: 10.1145/1772690.1772751]
59. Darling E, Shiffman D, Côté I, Drew J. The role of Twitter in the life cycle of a scientific publication. IEE 2013;6:32-43 [FREE Full text] [doi: 10.4033/iee.2013.6.6.f]
60. Huberman B, Romero DM, Wu F. Social networks that matter: Twitter under the microscope. FM 2008 Dec 20:1-9. [doi: 10.5210/fm.v14i1.2317]
61. Sola M, Sanchez-Quevedo C, Martin-Piedra MA, Carriel V, Garzon I, Chato-Astrain J, et al. Evaluation of the awareness of novel advanced therapies among family medicine residents in Spain. PLoS One 2019;14(4):e0214950 [FREE Full text] [doi: 10.1371/journal.pone.0214950] [Medline: 30943248]
62. Choo EK, Ranney ML, Chan TM, Trueger NS, Walsh AE, Tegtmeyer K, et al. Twitter as a tool for communication and knowledge exchange in academic medicine: A guide for skeptics and novices. Med Teach 2015 May;37(5):411-416. [doi: 10.3109/0142159X.2014.993371] [Medline: 25523012]
63. Gunn W. Social Signals Reflect Academic Impact. 2013. URL: https://www.niso.org/sites/default/files/stories/2017-08/IP_Gunn_Mendeley_isqv25no2.pdf [accessed 2021-11-02]
64. Glänzel W, Chi P. The big challenge of Scientometrics 2.0: exploring the broader impact of scientific research in public health. Scientometrics 2020 Apr 26;125(2):1011-1031. [doi: 10.1007/s11192-020-03473-x]
65. Bornmann L. Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics. Scientometrics 2015 Mar 15;103(3):1123-1144. [doi: 10.1007/s11192-015-1565-y]
66. O'Rourke C, Day AGE, Murray-Dunning C, Thanabalasundaram L, Cowan J, Stevanato L, et al. An allogeneic 'off the shelf' therapeutic strategy for peripheral nerve tissue engineering using clinical grade human neural stem cells. Sci Rep 2018 Feb 13;8(1):2951 [FREE Full text] [doi: 10.1038/s41598-018-20927-8] [Medline: 29440680]
67. Das S, Browne KD, Laimo FA, Maggiore JC, Hilman MC, Kaisaier H, et al. Pre-innervated tissue-engineered muscle promotes a pro-regenerative microenvironment following volumetric muscle loss. Commun Biol 2020 Jun 25;3(1):330 [FREE Full text] [doi: 10.1038/s42003-020-1056-4] [Medline: 32587337]
68. Murdoch CE, Scott CT. Stem cell tourism and the power of hope. Am J Bioeth 2010 May 10;10(5):16-23. [doi: 10.1080/15265161003728860] [Medline: 20461637]
69. Thelwall M, Wilson P. Mendeley readership altmetrics for medical articles: An analysis of 45 fields. J Assn Inf Sci Tec 2015 May 05;67(8):1962-1972. [doi: 10.1002/asi.23501]
70. Haustein S, Larivière V, Thelwall M, Amyot D, Peters I. Tweets vs. Mendeley readers: How do these two social media metrics differ? arXiv Preprint posted online on 2014. [FREE Full text] [doi: 10.1515/itit-2014-1048]
71. Mohammadi E, Thelwall M. Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. J Assn Inf Sci Tec 2014 Mar 12;65(8):1627-1638. [doi: 10.1002/asi.23071]
72. Haralabopoulos G, Anagnostopoulos I, Zeadally S. Lifespan and propagation of information in On-line Social Networks: A case study based on Reddit. Journal of Network and Computer Applications 2015 Oct;56:88-100. [doi: 10.1016/j.jnca.2015.06.006]
73. Berger J, Milkman KL. What Makes Online Content Viral? Journal of Marketing Research 2012 Apr 01;49(2):192-205. [doi: 10.1509/jmr.10.0353]
74. Pogorielov M. Tissue Engineering: Challenges and Selected Application. ATROA 2017 Nov 7;3(2):330-334. [doi: 10.15406/atroa.2017.03.00058]
75. Li X, Thelwall M, Giustini D. Validating online reference managers for scholarly impact measurement. Scientometrics 2011 Dec 21;91(2):461-471. [doi: 10.1007/s11192-011-0580-x]
76. Bik HM, Goldstein MC. An introduction to social media for scientists. PLoS Biol 2013;11(4):e1001535 [FREE Full text] [doi: 10.1371/journal.pbio.1001535] [Medline: 23630451]

77. Weigold M. Communicating Science. Science Communication 2016 Aug 18;23(2):164-193. [doi: 10.1177/1075547001023002005]
78. Thelwall M. Three practical field normalised alternative indicator formulae for research evaluation. Journal of Informetrics 2017 Feb;11(1):128-151. [doi: 10.1016/j.joi.2016.12.002]
79. Orduna-Malea E, Delgado LE. Demography of Altmetrics under the Light of Dimensions: Locations, Institutions, Journals, Disciplines and Funding Bodies in the Global Research Framework. Journal of Altmetrics 2019 Jun 18;2(1):1-18. [doi: 10.29024/joa.13]
80. Thelwall M. The Pros and Cons of the Use of Altmetrics in Research Assessment. Scholarly Assessment Reports 2020 May 12;2(1):1-9 [FREE Full text] [doi: 10.29024/sar.10]
81. Bardus M, El Rassi R, Chahrour M, Akl EW, Raslan AS, Meho LI, et al. The Use of Social Media to Increase the Impact of Health Research: Systematic Review. J Med Internet Res 2020 Jul 06;22(7):e15607 [FREE Full text] [doi: 10.2196/15607] [Medline: 32628113]
82. Costas R, Zahedi Z, Wouters P. Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. J Assn Inf Sci Tec 2014 Jul 28;66(10):2003-2019. [doi: 10.1002/asi.23309]
83. Kudlow P, Bissky Dziadyk D, Rutledge A, Shachak A, Eysenbach G. The citation advantage of promoted articles in a cross‐publisher distribution platform: A 12‐month randomized controlled trial. Journal of the Association for Information Science and Technology 2019 Dec 23;71(10):1257-1274. [doi: 10.1002/asi.24330]
84. Sinatra R, Deville P, Szell M, Wang D, Barabási A. A century of physics. Nature Phys 2015 Oct 01;11(10):791-796. [doi: 10.1038/nphys3494]
85. Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. EPI 2020 Jan 19;29(1):e290103. [doi: 10.3145/epi.2020.ene.03]

Abbreviations

TE: tissue engineering
WoS: Web of Science

Edited by R Kukafka; submitted 30.10.20; peer-reviewed by Z Ye, Y Bu; comments to author 21.01.21; revised version received 23.02.21; accepted 05.07.21; published 26.11.21.

Please cite as:
Santisteban-Espejo A, Martin-Piedra MA, Campos A, Moran-Sanchez J, Cobo MJ, Pacheco-Serrano AI, Moral-Munoz JA
Information and Scientific Impact of Advanced Therapies in the Age of Mass Media: Altmetrics-Based Analysis of Tissue Engineering
J Med Internet Res 2021;23(11):e25394
URL: https://www.jmir.org/2021/11/e25394
doi:10.2196/25394
PMID:34842548

Corrigenda and Addenda

Authorship Correction: International Changes in COVID-19 Clinical Trajectories Across 315 Hospitals and 6 Countries: Retrospective Cohort Study

Griffin M Weber[1*], MD, PhD; Harrison G Zhang[1*], HSDG; Sehi L'Yi[1*], PhD; Clara-Lea Bonzel[1], MSc; Chuan Hong[1], PhD; Paul Avillach[1], MD, PhD; Alba Gutiérrez-Sacristán[1], PhD; Nathan P Palmer[1], PhD; Amelia Li Min Tan[1], BSc, PhD; Xuan Wang[1], PhD; William Yuan[1], PhD; Nils Gehlenborg[1], PhD; Anna Alloni[2], PhD; Danilo F Amendola[3], MSc; Antonio Bellasi[4], MD, PhD; Riccardo Bellazzi[5], MS, PhD; Michele Beraghi[6], MS; Mauro Bucalo[2], MS; Luca Chiovato[7], MD, PhD; Kelly Cho[8], PhD; Arianna Dagliati[9], MS, PhD; Hossein Estiri[10], PhD; Robert W Follett[11], BS; Noelia García Barrio[12], MS; David A Hanauer[13], MS, MD; Darren W Henderson[14], BS; Yuk-Lam Ho[8], MPH; John H Holmes[15,16], MS, PhD; Meghan R Hutch[17], BS; Ramakanth Kavuluru[18], PhD; Katie Kirchoff[19], MSHI; Jeffrey G Klann[10], MEng, PhD; Ashok K Krishnamurthy[20], PhD; Trang T Le[15], PhD; Molei Liu[21], MSc; Ne Hooi Will Loh[22], MBBS; Sara Lozano-Zahonero[23], PhD; Yuan Luo[17], PhD; Sarah Maidlow[24], AA; Adeline Makoudjou[23], MD; Alberto Malovini[25], PhD; Marcelo Roberto Martins[26], MSc; Bertrand Moal[27], MD, PhD; Michele Morris[28], BA; Danielle L Mowery[15], PhD; Shawn N Murphy[29], MD, PhD; Antoine Neuraz[30], MD, PhD; Kee Yuan Ngiam[31], MBBS, MRCS, MMed; Marina P Okoshi[32], MD, PhD; Gilbert S Omenn[33], MD, PhD; Lav P Patel[34], MS; Miguel Pedrera Jiménez[12], MS; Robson A Prudente[32], PhD; Malarkodi Jebathilagam Samayamuthu[28], MD; Fernando J Sanz Vidorreta[11], BS; Emily R Schriver[35], MS; Petra Schubert[8], MPH; Pablo Serrano Balazote[12], MS, MD; Byorn WL Tan[36], MBBS; Suzana E Tanni[32], MD, PhD; Valentina Tibollo[25], MS; Shyam Visweswaran[28], MD, PhD; Kavishwar B Wagholikar[10], MBBS, PhD; Zongqi Xia[37], MD, PhD; Daniela Zöller[23], PhD; The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)[39]; Isaac S Kohane[1], MD, PhD; Tianxi Cai[1*], ScD[‡]; Andrew M South[38*], MS, MD; Gabriel A Brat[1*], MPH, MD

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[2]BIOMERIS (BIOMedical Research Informatics Solutions), Pavia, Italy

[3]Clinical Research Unit, Botucatu Medical School, São Paulo State University, Botucatu, Brazil

[4]Division of Nephrology, Department of Medicine, Ente Ospedaliero Cantonale, Lugano, Switzerland

[5]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[6]Information Technology Department, Azienda Socio-Sanitaria Territoriale di Pavia, Pavia, Italy

[7]Unit of Internal Medicine and Endocrinology, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy

[8]Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Boston Healthcare System, Boston, MA, United States

[9]Department of Electrical Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[10]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

[11]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[12]Health Informatics, Hospital Universitario 12 de Octubre, Madrid, Spain

[13]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States

[14]Department of Biomedical Informatics, University of Kentucky, Lexington, KY, United States

[15]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[16]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[17]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[18]Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, United States

[19]Medical University of South Carolina, Charleston, SC, United States

[20]Department of Computer Science, Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[21]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[22]Department of Anaesthesia, National University Health System, Singapore, Singapore

[23]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

[24]Michigan Institute for Clinical & Health Research Informatics, University of Michigan, Ann Arbor, MI, United States

[25]Laboratory of Informatics and Systems Engineering for Clinical Research, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy
[26]Clinical Hospital of Botucatu Medical School, São Paulo State University, Botucatu, Brazil
[27]Informatique et archivistique médicales unit, Bordeaux University Hospital, Bordeaux, France
[28]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States
[29]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States
[30]Department of Biomedical Informatics, Hôpital Necker-Enfants Malade, Assistance Publique Hôpitaux de Paris, University of Paris, Paris, France
[31]Department of Biomedical Informatics, Institute for Digital Medicine, National University Health System, Singapore, Singapore
[32]Internal Medicine Department, Botucatu Medical School, São Paulo State University, Botucatu, Brazil
[33]Department of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics, and Public Health, University of Michigan, Ann Arbor, MI, United States
[34]Division of Medical Informatics, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS, United States
[35]Data Analytics Center, University of Pennsylvania Health System, Philadelphia, PA, United States
[36]Department of Medicine, National University Health System, Singapore, Singapore
[37]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States
[38]Section of Nephrology, Department of Pediatrics, Brenner Children's Hospital, Wake Forest School of Medicine, Winston Salem, NC, United States
[39]see Authors' Contributions
[‡]The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)
[*]these authors contributed equally

Corresponding Author:
Tianxi Cai, ScD
Department of Biomedical Informatics
Harvard Medical School
10 Shattuck St
Boston, MA, 02115
United States
Phone: 1 6175488520
Email: tcai@hsph.harvard.edu

Related Article:

Correction of: https://www.jmir.org/2021/10/e31400

(J Med Internet Res 2021;23(11):e34625) doi:10.2196/34625

In "International Changes in COVID-19 Clinical Trajectories Across 315 Hospitals and 6 Countries: Retrospective Cohort Study" (J Med Internet Res 2021 Oct 11;23(10):e31400), two errors were noted.

In the originally published paper, equal contribution of the last three authors was not noted. This has been corrected to add a note of equal contribution to the last three authors, as well as to add a statement to the Authors' Contributions section denoting the nature of equal contributions.

In the originally published paper, an equal contribution footnote was applied to authors Griffin M Weber, Harrison G Zhang, and Sehi L'Yi.

In the corrected paper, the equal contribution footnote has been applied to authors Griffin M Weber, Harrison G Zhang, Sehi L'Yi, Tianxi Cai, Andrew M South, and Gabriel A Brat.

The Authors' Contribution section has been updated to include the following statement:

> *These authors contributed equally: GMW, HGZ, SL. These authors jointly supervised the work: TC, AMS, GAB.*

The correction will appear in the online version of the paper on the JMIR Publications website on November 30, 2021, together with the publication of this correction notice.

Submitted 02.11.21; this is a non–peer-reviewed article;accepted 10.11.21; published 30.11.21.

Please cite as:
Weber GM, Zhang HG, L'Yi S, Bonzel CL, Hong C, Avillach P, Gutiérrez-Sacristán A, Palmer NP, Tan ALM, Wang X, Yuan W, Gehlenborg N, Alloni A, Amendola DF, Bellasi A, Bellazzi R, Beraghi M, Bucalo M, Chiovato L, Cho K, Dagliati A, Estiri H, Follett RW, García Barrio N, Hanauer DA, Henderson DW, Ho YL, Holmes JH, Hutch MR, Kavuluru R, Kirchoff K, Klann JG, Krishnamurthy AK, Le TT, Liu M, Loh NHW, Lozano-Zahonero S, Luo Y, Maidlow S, Makoudjou A, Malovini A, Martins MR, Moal B, Morris M, Mowery DL, Murphy SN, Neuraz A, Ngiam KY, Okoshi MP, Omenn GS, Patel LP, Pedrera Jiménez M, Prudente RA, Samayamuthu MJ, Sanz Vidorreta FJ, Schriver ER, Schubert P, Serrano Balazote P, Tan BWL, Tanni SE, Tibollo V, Visweswaran S, Wagholikar KB, Xia Z, Zöller D, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Kohane IS, Cai T, South AM, Brat GA
Authorship Correction: International Changes in COVID-19 Clinical Trajectories Across 315 Hospitals and 6 Countries: Retrospective Cohort Study
J Med Internet Res 2021;23(11):e34625
URL: https://www.jmir.org/2021/11/e34625
doi:10.2196/34625
PMID:34889759

Original Paper

Utility of a Telephone Triage Hotline in Response to the COVID-19 Pandemic: Longitudinal Observational Study

Benjamin A Y Cher[1*], MS; Eric A Wilson[1*], BS; Alexa M Pinsky[1], BS; Ryan F Townshend[1], BS; Ann V Wolski[1], BS; Michael Broderick[1], BS; Allison M Milen[1], BS; Audrey Lau[1], BS; Amrit Singh[1], BS; Sandro K Cinti[1,2], MD; Carl G Engelke[1*], BS; Anjan K Saha[3*], MD, PhD

[1]Michigan Medicine, University of Michigan, Ann Arbor, MI, United States

[2]Department of Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, MI, United States

[3]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Anjan K Saha, MD, PhD
Department of Anesthesiology
Columbia University Irving Medical Center
622 West 168th Street
10032
New York, NY
United States
Phone: 1 (703) 785 3087
Email: aks9032@nyp.org

Abstract

Background: During the initial months of the COVID-19 pandemic, rapidly rising disease prevalence in the United States created a demand for patient-facing information exchanges that addressed questions and concerns about the disease. One approach to managing increased patient volumes during a pandemic involves the implementation of telephone-based triage systems. During a pandemic, telephone triage hotlines can be employed in innovative ways to conserve medical resources and offer useful population-level data about disease symptomatology and risk factor profiles.

Objective: The aim of this study is to describe and evaluate the COVID-19 telephone triage hotline used by a large academic medical center in the midwestern United States.

Methods: Michigan Medicine established a telephone hotline to triage inbound patient calls related to COVID-19. For calls received between March 24, 2020, and May 5, 2020, we described total call volume, data reported by callers including COVID-19 risk factors and symptomatology, and distribution of callers to triage algorithm endpoints. We also described symptomatology reported by callers who were directed to the institutional patient portal (online medical visit questionnaire).

Results: A total of 3929 calls (average 91 calls per day) were received by the call center during the study period. The maximum total number of daily calls peaked at 211 on March 24, 2020. Call volumes were the highest from 6 AM to 11 AM and during evening hours. Callers were most often directed to the online patient portal (1654/3929, 42%), nursing hotlines (1338/3929, 34%), or employee health services (709/3929, 18%). Cough (126/370 of callers, 34%), shortness of breath (101/370, 27%), upper respiratory infection (28/111, 25%), and fever (89/370, 24%) were the most commonly reported symptoms. Immunocompromised state (23/370, 6%) and age >65 years (18/370, 5%) were the most commonly reported risk factors.

Conclusions: The triage algorithm successfully diverted low-risk patients to suitable algorithm endpoints, while directing high-risk patients onward for immediate assessment. Data collected from hotline calls also enhanced knowledge of symptoms and risk factors that typified community members, demonstrating that pandemic hotlines can aid in the clinical characterization of novel diseases.

(J Med Internet Res 2021;23(11):e28105) doi:10.2196/28105

KEYWORDS

triage; telephone; COVID-19; utility; telemedicine; telehealth; patient information; concern; implementation; innovation; hospital

Introduction

COVID-19 emerged in December 2019 as a highly contagious viral pneumonia caused by SARS-CoV-2 [1]. The disease is transmitted through aerosolized respiratory droplets and was classified as a pandemic by the World Health Organization on March 11, 2020, after rapidly crossing intercontinental and transoceanic boundaries over three months [2]. Symptoms include fever, cough, and dyspnea; unmanaged disease can result in multiorgan failure and possibly death [3,4]. The global response to COVID-19 has focused on medically managing the disease, optimizing hospital utilization, and reducing viral transmission [5].

The rising prevalence of COVID-19 in the United States [6] created a demand for patient-facing information exchanges that addressed questions and concerns about the disease. Local efforts designed to contextualize national facts and figures were critical for layering population-specific considerations over nationally accepted guidelines. As the disease began to spread, acute care settings experienced a surge in patients seeking treatment and advice, stressing already overburdened health care systems and creating infection control challenges in overcrowded waiting rooms [7,8]. Methods aimed at managing this sudden spike in health care utilization required adherence to social distancing recommendations to prevent viral transmission while simultaneously alleviating the added systemic congestion attributed to COVID-19.

One approach to managing increased patient volumes during a pandemic involves the implementation of telephone-based triage systems. Triage and evaluation phone lines established during previous pandemics were found to be cost-effective methods for prioritizing acute care resources for symptomatic patients with risk factors for serious disease in overburdened hospital systems [9-11]. A study of infectious disease hotlines implemented following natural disasters showed that these systems effectively aid patient evaluation and prevent disease outbreaks in vulnerable populations [12]. Furthermore, disease triage hotlines serve as effective information waypoints for collecting longitudinal patient data while relieving congestion that would otherwise exacerbate disease propagation [13]. However, a lack of consensus on a universal approach to triage due to variation in disease presentation across institutions and geographic areas prevents scalable implementation of triage hotlines [14-16]. Approaches that optimize resource allocation by accounting for variation in disease presentation thus present compelling value propositions for health systems worldwide to address ongoing and future pandemics.

To this end, we aimed to evaluate the COVID-19 triage hotline established by Michigan Medicine, the academic medical center affiliated with the University of Michigan, which serves the population of Southeast Michigan and surrounding areas. We characterized the call center's triage algorithm and described the distribution of all calls fielded over a 6-week interval during the initial surge of the pandemic. We then contextualized hourly and daily call volumes against publicly available internet search trends concerning COVID-19–related inquiries. Lastly, we tabulated symptomatology and risk factor profiles for patients who accessed call center services during the 6-week period to provide perspective on local disease burden. Our results illustrate the utility of telephone-based triage and evaluation hotlines in optimizing health care utilization and surveying disease burden during times of crisis that impact human health.

Methods

COVID-19 Call Center Hotline and Triage Algorithm

Michigan Medicine established a telephone hotline to triage inbound patient calls related to COVID-19. Operations commenced on March 16, 2020, and accepted calls daily from 6 AM to 12 AM. The hotline was staffed both in person and remotely by call center agents, medical assistants, patient service representatives, and medical student volunteers (herein termed as "agents"). The call center used Aspect Call Center software (Aspect) to receive and manage calls. Call center agents used an algorithm developed by the Michigan Medicine Ambulatory Care Administration to triage incoming calls with screening questions and defined endpoints. Screening questions were updated over the duration of the study to accommodate the changing health landscape of the pandemic. Symptom screening criteria were initially limited to fever, cough, and dyspnea. Criteria were expanded to include upper respiratory symptoms (runny nose and congestion), muscle aches, loss of sense of smell, and diarrhea on April 20, 2020. A positive symptom screen was defined as having at least two symptoms from this list. Risk factor criteria initially included age over 65 years, compromised immune system, chronic respiratory illness, presence of a health care worker in the household, hemodialysis patient, resident of a skilled nursing or long-term care facility, and/or living in communal housing. Risk factor criteria were expanded to include primary caretaker for a vulnerable individual and/or close contact of an individual who tested positive for COVID-19 on April 20, 2020. A positive risk factor screen was defined as having at least one risk factor from this list.

Data Sources

We used two surveys to collect data for this analysis. The first was created in Qualtrics, an electronic survey platform. Call center workers filled out a Qualtrics survey when they handled each call to the call center. From these results, we extracted call volume, patient employment status (Michigan Medicine—and later, University of Michigan—employee versus nonemployee), patient status (Michigan Medicine versus non–Michigan Medicine patient), pediatric or adult, symptom profile, and risk factor data**.** The triage survey directed callers toward various algorithm endpoints: Occupational Health Services, non–Michigan Medicine primary care provider or the Centers for Disease Control and Prevention website, patient portal e-visit questionnaire, and adult or pediatric nursing hotlines. Partially completed Qualtrics survey responses were discarded from downstream analysis. The second survey was completed by a subset of call center workers who were volunteers during evening shifts. This survey was created in Google Forms, an additional online survey platform. The Google Forms survey collected additional data about symptomatology and risk factors that were not included in the Qualtrics survey. Qualtrics and

Google Forms survey data were merged to produce a composite data set at the end of the study period. We used an additional data source, the Aspect Unified IP System, which automatically logged all telephone calls to our institution's call center hotline. We used this source to cross-check data on call volumes extracted from the Qualtrics and Google surveys.

To contextualize how call volume was related to public interest in the pandemic and disease prevalence, we accessed publicly available data through two sources: (1) Google Trends and (2) the State of Michigan online coronavirus dashboard [17]. We compiled regional trends in searches for the terms "COVID," "COVID-19," and "Coronavirus" during the study period. We also compiled search trends for words that mirrored the symptoms in our screening algorithm: "Fever," "Cough," "Short of Breath," "Shortness of Breath," and "Trouble Breathing." Data provided by Google Trends are normalized to a maximum of 100 for the frequency of searches between March 24, 2020, and May 5, 2020, for the state of Michigan and the city of Detroit. Data provided by the State of Michigan online coronavirus dashboard were presented as 3-day total confirmed cases and deaths from COVID-19 in Washtenaw county during the same time period.

In tandem with the creation of a call center to assess patients, Michigan Medicine created a "Cough, Flu, and COVID-19–like Symptoms" patient portal e-visit questionnaire. Patients calling the triage hotline whose symptoms or risk factor profile did not meet criteria to be transferred to the nursing line were directed to complete this e-visit questionnaire in their patient portal. The questionnaire assessed symptom frequencies, symptom onset and duration, and epidemiologic and demographic risk factors. Once completed, a provider would reach out to the patient to determine appropriate next steps in care.

Data Analysis

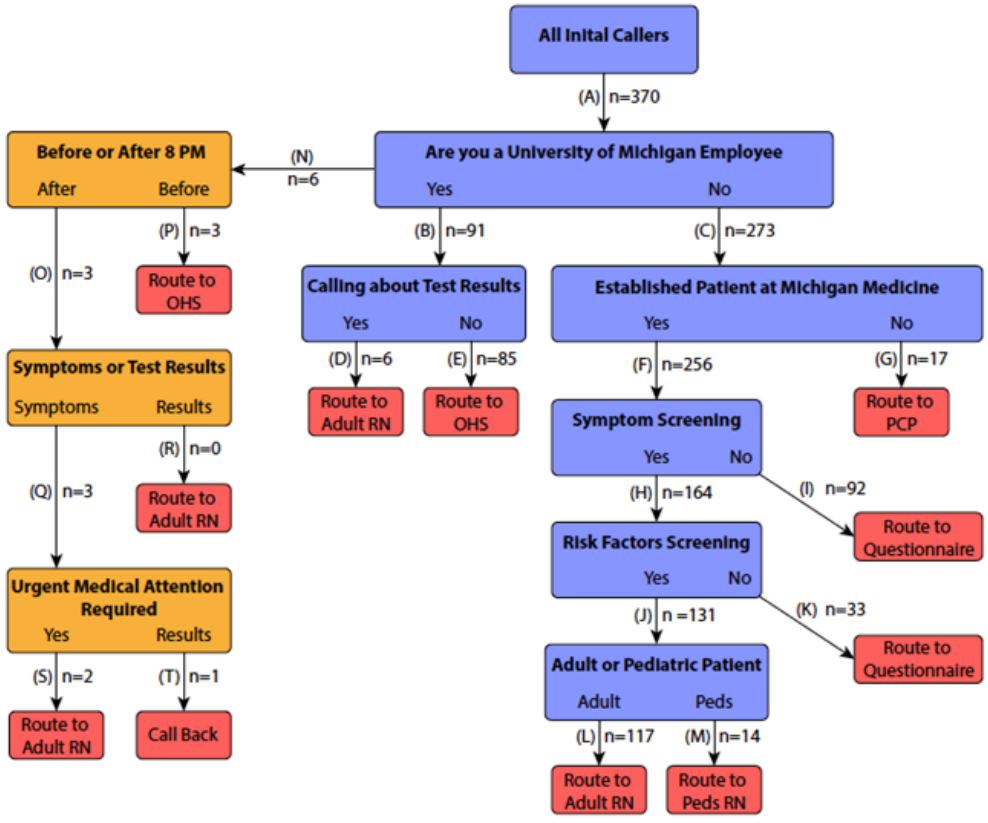
Data were collected between March 24, 2020 (8 days after the triage hotline began operation) and May 5, 2020. Total call volume and the total number of calls directed to each triage algorithm endpoint by date and time of day were calculated to assess hotline workload. Symptom frequency was subsequently overlaid on total call volume for the call center and the total number of calls in which patients screened positive for symptoms. Google Forms survey data and the "Cough, Flu, and COVID-19–like Symptoms" patient portal questionnaire responses were analyzed to determine the frequency of patient-reported symptoms and epidemiologic risk factors among both hotline callers and online portal users. Data representation and statistical analysis was completed using Prism (version 8.4.2; GraphPad). The study was reviewed and approved by the local institutional review board (HUM00179879).

Results

Call Volumes and Algorithm Endpoints

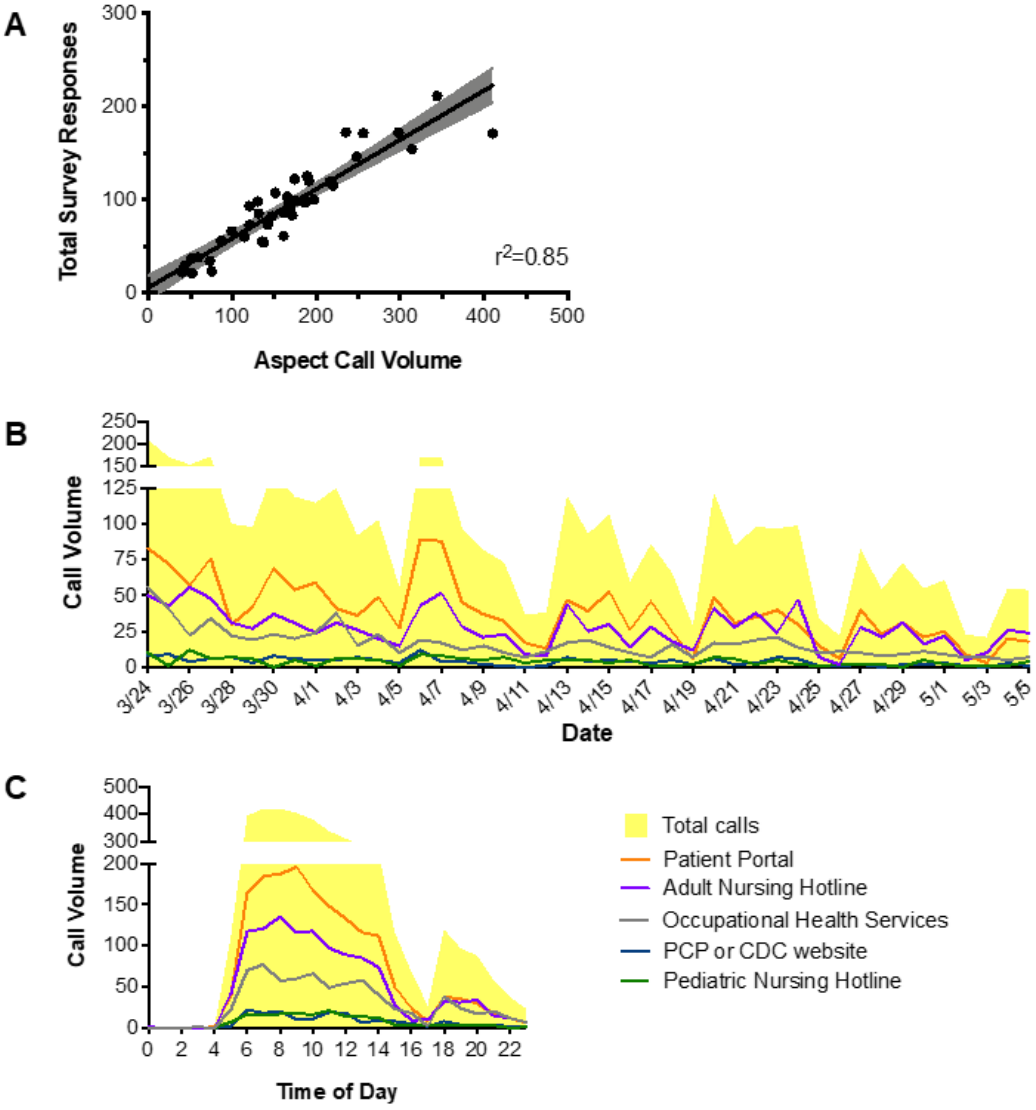
Upon establishing a staffed triage hotline, calls were processed through the screening algorithm, which directed patients to specific algorithm endpoints (Figure 1).

Figure 1. Detailed workflow of the Michigan Medicine COVID-19 Hotline. Outline of triage algorithm, including initial input/questions (blue), endpoints (red), and an alternative pathway added on May 1, 2020 (orange). The numbers of callers routed through each pathway are shown at unique junctions in the algorithm. OHS: Occupational Health Services; PCP: primary care provider; RN: registered nurse.



To establish that our survey data provided an accurate representation of our call volume, we compared daily call volume to the number of calls routed to the hotline number through Aspect Unified IP, which automatically logged all calls to the call center hotline (Figure 2A). Although the Google and Qualtrics survey response rate was approximately 50% of Aspect system forwarding, the linear relationship between call volume as measured by the surveys and the Aspect system (r^2=0.85) suggests that survey responses provide a reasonable representation of call volume received by the triage hotline.

Figure 2. Linear regression analysis of survey responses versus number of incoming calls and daily/hourly call volumes. (A) The number of unique Qualtrics and Google Forms survey responses collected during the study period had a strong positive correlation with COVID-19 Hotline call volumes recorded by Aspect software during the same time frame (Pearson coefficient, r^2=0.85). (B) Daily and (C) hourly call volumes obtained from combined Qualtrics and Google Forms survey data, with total calls received graphed as a stacked line plot (yellow) with subcategories of triage algorithm endpoints overlaid as line graphs. Orange: routed to patient portal; purple: routed to adult nursing; grey: routed to OHS; blue: routed to PCP or Centers for Disease Control and Prevention website; green: routed to pediatric nursing. OHS: Occupational Health Services; PCP: primary care provider.

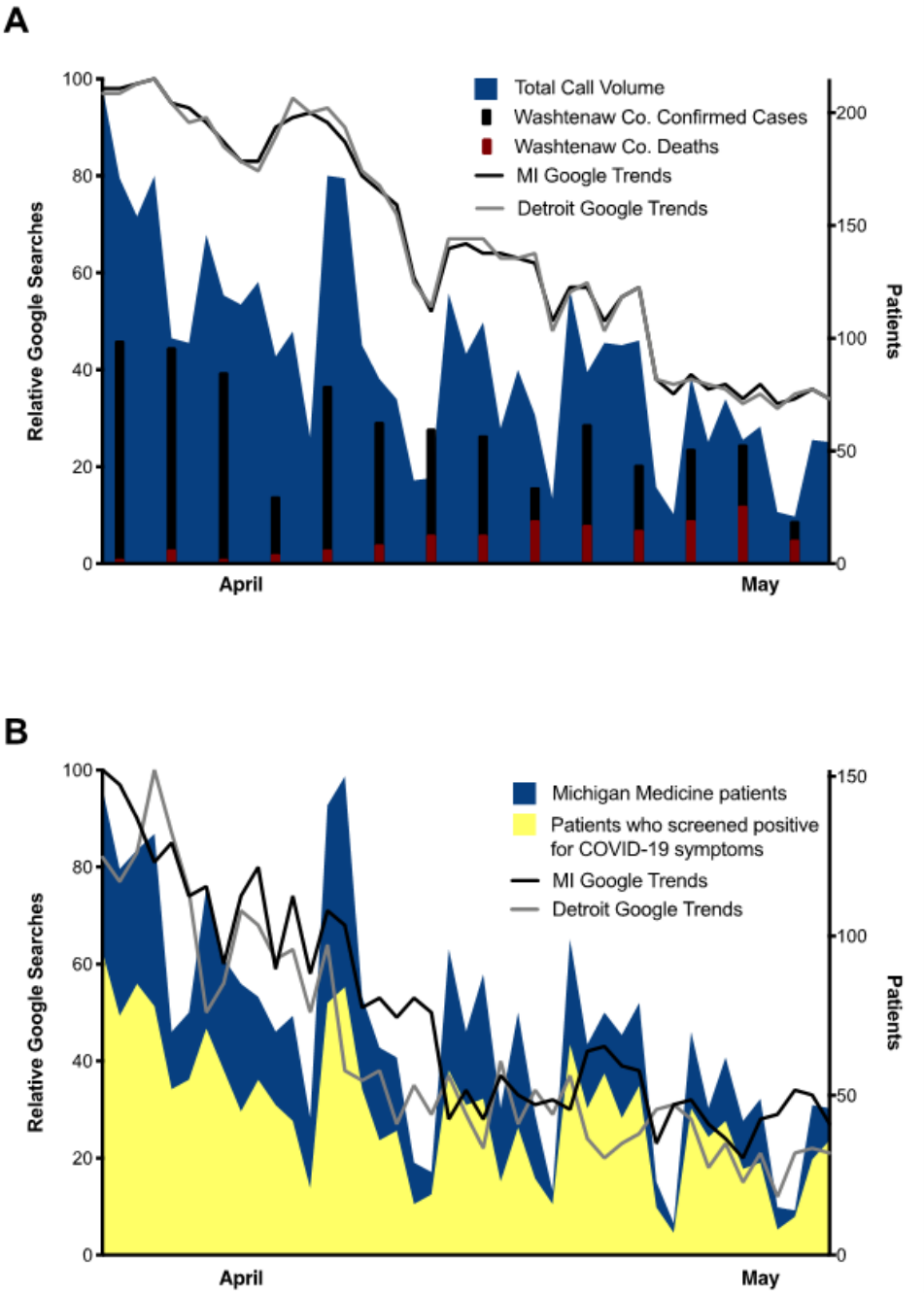


A total of 3929 calls (average 91 calls per day) were received by the call center during the study period. Overall, call volume gradually declined over time but showed substantial variation between days throughout the study period (Figure 2B). The distribution of calls between algorithm endpoints was consistent through the study period. Callers were most often directed to the institutional patient portal (1654/3929, 42%), nursing hotlines (1338/3929, 34%), or Occupational Health Services (709/3929, 18%). Of the 1338 calls directed to the nursing hotline, 1164 (1164/1338, 87%) were forwarded to the adult nursing hotline and 174 (174/1338, 13%) were sent to the pediatric nursing hotline. The total number of calls peaked at 211/day on March 24, 2020, and fell to 54/day by May 5, 2020, when data collection ended. Incoming call volumes were highest between 6 AM and 11 AM and steadily decreased throughout the remainder of the day with the exception of a brief resurgence between 6 PM and 12 AM (Figure 2C). Additionally, hourly call volumes were lower on the weekends than during weekdays, but volumes were similarly distributed throughout the day (Figure S1 in Multimedia Appendix 1).

We were interested to see whether call volume and the fraction of callers experiencing symptoms was reflective of the public interest in the pandemic and locoregional case volume. Google Trends frequencies for COVID-19–related searches for the state

of Michigan and the Detroit area were highest in late March and gradually decreased until the end of data collection in early May (Figure 3A). Confirmed cases for Washtenaw County and total triage hotline call volume followed a similar overall trend, though both demonstrated lulls followed by sudden spikes in early and mid-April. Searches related to the most common symptoms of COVID-19 in both Michigan and Detroit peaked in late March and early April, mirroring the peak in call volume and callers screening positive for symptoms (Figure 3B).

Figure 3. Google Search and hotline call volume trends. (A) Total call volume derived from combined Qualtrics and Google Forms survey data graphed as a stacked line plot (blue), relative frequencies of COVID-19–related Google Search results for the state of Michigan (MI, black) and Detroit area (grey) overlaid as line plots, and 3-day total confirmed cases (black) and deaths (red) from COVID-19 in Washtenaw county represented as bars. (B) Total number of callers who were Michigan Medicine patients (blue stacked line plot) and those who screened positive for symptoms (yellow stacked line plot) overlaid with relative frequencies of Google Search results pertaining to COVID-19 symptoms in the state of Michigan (black) and Detroit area (grey) overlaid as line plots. For panels (A) and (B), the left y-axis represents the normalized range (0-100) of Google Trends results, and call volumes and patient numbers are represented on the right y-axis.

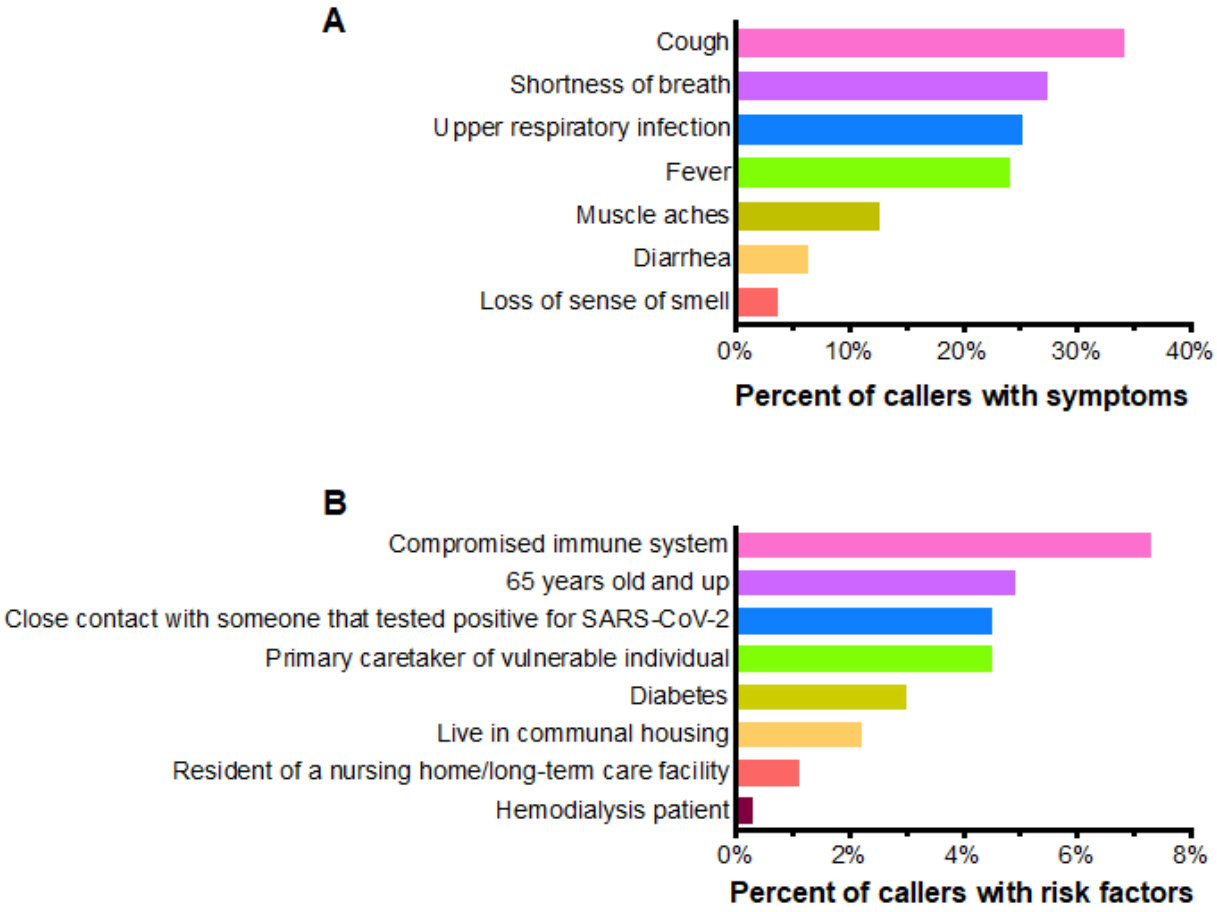


Caller Symptomatology and Risk Factors

We next analyzed Google Forms survey data to determine the frequency of patient-reported symptoms and the distribution of epidemiologic risk factors among callers. These represent the subset of callers who were Michigan Medicine patients and contacted the call center between 6 PM and 12 AM. Cough (126/370, 34% of callers), shortness of breath (101/370, 27%), upper respiratory infection (28/111, 25%), and fever (89/370, 24%) were the most commonly reported symptoms (Figure 4A).

A minority of callers reported muscle aches, diarrhea, or loss of sense of smell (14/111, 13%; 7/111, 6%; and 4/111, 4%, respectively). Among the risk factors assessed, immunocompromised state (23/370, 7%), age >65 years (18/370, 5%), primary caretaker of a vulnerable individual (5/370, 4.5%), and close contact with a person known to have tested positive for SARS-CoV-2 (5/370, 4.5%) were the most frequently reported by hotline users (Figure 4B).

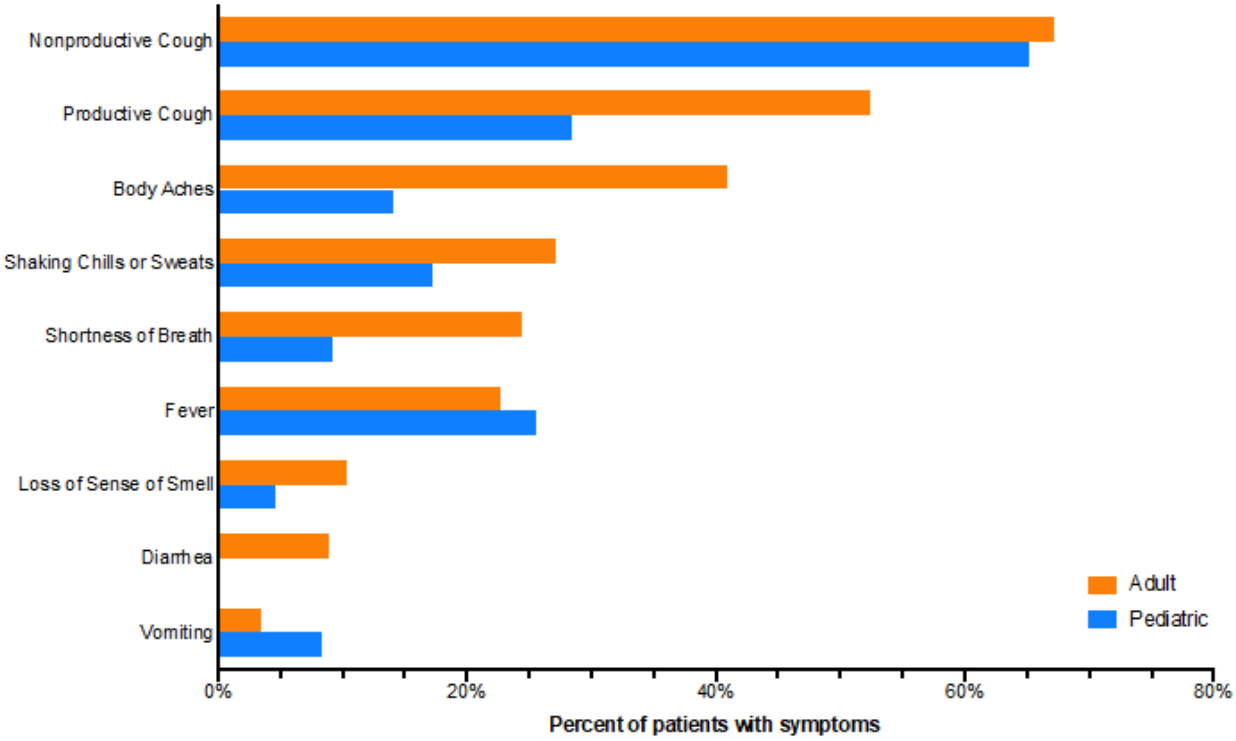
Figure 4. Symptoms and risk factors reported by hotline callers. Percent of patient-reported (A) symptoms and (B) risk factors obtained from Google Forms survey data. For "Cough," "Shortness of breath," "Fever," "Compromised immune system," "65 years old and up," "Hemodialysis patient," "Diabetes," "Resident of a nursing home/long-term care facility," and "Living in communal housing," n=370, while for "Upper respiratory infection," "Muscle aches," "Diarrhea," "Loss of sense of smell," "Close contact with someone that tested positive for SARS-CoV-2," and "Primary caretaker of a vulnerable individual," n=111, as these were added to the screening algorithm at a later study timepoint.



Patient Portal Symptomatology and Risk Factors

Further characterization of patient-reported symptoms and epidemiologic risk factors was achieved through “Cough, Flu, and COVID-19–like Symptoms” questionnaire responses. Questionnaire data revealed that nonproductive or productive cough were the most frequently reported symptoms among both adult (nonproductive: 1280/1907, 67%; productive: 998/1907, 52%) and pediatric (nonproductive: 148/227, 65%; productive: 64/227, 28%) patients (Figure 5). Muscle aches were the next most common symptom reported by adults (780/1907, 41%), while pediatric patients were more likely to experience fever (58/227, 26%). Adult patients were more likely to experience dyspnea than pediatric patients (465/1907, 24% versus 21/227, 9%) and loss of sense of smell (198/1907, 10% versus 10/227, 5%). Gastrointestinal symptoms (vomiting, diarrhea) were reported by <10% of adult and pediatric patients. Regarding clinical time course, both adult and pediatric patients were more likely to report a slowly progressive disease course (67%, 1393/2081 of adults and 61%, 118/194 of pediatric patients) than acute onset of illness (33%, 688/2081 of adults and 39%, 76/194 of pediatric patients). Health care worker (12%, 259/2132 of adults) and household contact of a health care worker (15%, 30/196 of pediatric patients) were the most frequently reported risk factors (Figure S2A,B in Multimedia Appendix 1). Of adult patients, 7% (142/2132) indicated that they had a compromised immune system compared to 2% (4/196) of pediatric patients.

Figure 5. "Cough, Flu, and COVID-19–like Symptoms" e-visit questionnaire symptom profiles for adult and pediatric patients. (A) Percent of reported symptoms among adult (orange, n=1907) and pediatric (blue, n=227) patients who completed the e-visit questionnaire within the study time frame.



Discussion

Principal Findings

In this study, we illustrated an approach to rapid development and implementation of a telephone-based triage system during the COVID-19 pandemic. Although call volumes fluctuated considerably throughout the study period, we observed a gradual decline in total calls received over time. We noted that referral to the patient portal or transfer to the nursing hotline for further assessment constituted more than 75% of triage decisions. Google Trends data for COVID-19–related terminology and symptoms in Michigan and the Detroit metropolitan area decreased from late March to early May 2020, mirroring the declining trend in call volumes noted at the call center. Callers most frequently reported the following symptoms: cough, shortness of breath, upper respiratory infection, and fever. In addition, the following risk factors for infection were most frequently reported: immunocompromised, age greater than 65 years, and exposure to another individual who tested positive for COVID-19. Analysis of patient questionnaire data revealed that cough was the most frequently reported symptom by adult and pediatric patients, but the frequency of symptoms thereafter differed between these groups. Together, these results demonstrate how patient hotlines and triage systems can be used to properly allocate care and broaden our understanding of disease characteristics during public health crises.

Responding effectively to a global pandemic necessitates judicious use of resources and regular assessment of disease burden. Tools with this dual capability are positioned to meaningfully contribute to pandemic response efforts. Specifically, triage and evaluation phone services are well-equipped to manage and survey patient populations when deployed by health systems during pandemics. Characterization of the Michigan Medicine COVID-19 Hotline validated this notion. The call center successfully directed call traffic away from the resource-intensive nursing endpoints while capturing population-level statistics on disease symptomatology and prevalence. This was accomplished by direct triage and surveillance by call center agents as well as questionnaire-based surveillance through a pre-existing online portal. Only 17% of the total call volume was directed to the nursing hotlines, while 21% of calls were directed to the online patient portal. Cough was identified as the most common symptom by triage agents as well as questionnaires, coinciding well with prior reports on COVID-19 symptomatology [1,5]. Taken together, the Michigan Medicine COVID-19 Hotline effectively triaged patients seeking advice and care during the COVID-19 pandemic while facilitating characterization of local disease burden.

Prior triage hotlines have provided population-level statistics about disease symptomatology and prevalence while directing resource allocation [13]. During the H1N1 pandemic of 2009, triage lines were developed to facilitate rapid diversion of call traffic from upper-level providers [9,11,13] while providing reassurance to patients with less concerning symptoms [11]. In addition, total call volume may itself be a useful marker for informing appropriate resource management. The tight concordance observed between regional Google Trends search term frequencies and total call volume is in line with the findings of previous studies comparing search trends to disease incidence [18,19] and supports the notion that call center traffic can operate as a barometer for community concerns surrounding COVID-19.

Designing the triage algorithm to leverage an online patient portal not only reduced nursing utilization but also unlocked local care capacity without compromising social distancing measures designed to reduce viral transmission. Previous work has demonstrated that call centers can also benefit a community beyond their primary role as a triage service. For instance, South Australia's COVID-19 Relief Call Centre assisted callers with food insecurity and accessing medical appointments and welfare checks [20]. Call centers additionally provide alternative employment opportunities during pandemic shutdowns for both health care workers and non–health care workers. As an example, nurses and medical assistants on furlough at Michigan Medicine were redeployed as call center agents. In New York City, customer service agents at a travel management firm were trained and deployed as COVID-19 test schedulers [21]. Call centers have also been an effective means of reaching underserved patients in rural areas [22].

Nevertheless, a number of operational challenges place constraints on the value of call centers. Hotline staff do not perform physical exams, which are often necessary to properly evaluate the cardiopulmonary complaints that are characteristic of COVID-19. Maintaining social distancing standards and properly sanitizing shared workspaces can also pose a challenge to call center operations during pandemics [23]. To circumvent these challenges, hotlines based on yes/no algorithms can be supplemented by online self-triage systems [24]. Social distancing practices can be maintained by adopting work from home strategies, as was done by a call center in England that transitioned almost 1000 employees to remote work at the outset of the pandemic [25]. Although imbalances between health care supply and demand are likely to arise during future pandemics or subsequent waves of the COVID-19 pandemic, call center hotlines can be used to reduce the burden placed on strained medical systems.

Limitations

A number of operational constraints placed limitations on the scope of our study. First, the Google Forms survey designed to screen symptomatology and risk factors was only completed for evening callers. Specific symptoms or risk factors associated with the time frame in which calls were received may have confounded our results. Second, due to the structure of the triage algorithm, we were only able to present symptomatology and risk factor data for a small subset of patients. Further, since the triage algorithm screened callers as positive if they had two symptoms or one risk factor, it is possible that additional symptoms or risk factors were not captured once a caller had screened positive. Third, not all callers had questions or requests that fit perfectly into the triage algorithm. For these calls, call center workers may have skipped the online survey (leading to inaccurate measurements of call volume) or filled out the survey with their best approximation of the caller's needs. Fourth, due to the deidentified nature of our data sets, we were unable to measure caller satisfaction. This would be a useful topic for further research. Finally, this study presents data from one institution, limiting the ability of our data to generate population-level inferences.

Conclusions

In summary, our study describes the development and implementation of a COVID-19 patient triage hotline in a single health care system. The triage algorithm successfully diverted low-risk patients to suitable algorithm endpoints, while directing high-risk patients onward for immediate assessment. Data collected from hotline calls also enhanced our knowledge of typical symptoms and risk factors among community members, demonstrating that pandemic hotlines can aid in the clinical characterization of novel diseases. Although future innovation in the areas of triage algorithm design and remote work capabilities can certainly improve the operation of future pandemic hotlines, our work provides critical insight into the role that telephone-based triage systems play in facilitating health care delivery in times of crisis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[DOCX File , 83 KB - jmir_v23i11e28105_app1.docx]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
2. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
3. Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington state. JAMA 2020 Apr 28;323(16):1612-1614 [FREE Full text] [doi: 10.1001/jama.2020.4326] [Medline: 32191259]

4. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Wu Y, Chen C, Chan Y. The outbreak of COVID-19. Journal of the Chinese Medical Association 2020;83(3):217-220. [doi: 10.1097/jcma.0000000000000270]
6. CDC COVID-19 Response Team. Geographic Differences in COVID-19 Cases, Deaths, and Incidence - United States, February 12-April 7, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):465-471 [FREE Full text] [doi: 10.15585/mmwr.mm6915e4] [Medline: 32298250]
7. Carenzo L, Costantini E, Greco M, Barra FL, Rendiniello V, Mainetti M, et al. Hospital surge capacity in a tertiary emergency referral centre during the COVID-19 outbreak in Italy. Anaesthesia 2020 Jul 22;75(7):928-934 [FREE Full text] [doi: 10.1111/anae.15072] [Medline: 32246838]
8. Kearsley R, Duffy CC. The COVID-19 information pandemic: how have we managed the surge? Anaesthesia 2020 Aug 02;75(8):993-996 [FREE Full text] [doi: 10.1111/anae.15121] [Medline: 32406517]
9. Spaulding AB, Radi D, Macleod H, Lynfield R, Larson M, Hyduke T, et al. Satisfaction and public health cost of a statewide influenza nurse triage line in response to pandemic H1N1 influenza. PLoS One 2013 Jan 15;8(1):e50492 [FREE Full text] [doi: 10.1371/journal.pone.0050492] [Medline: 23335953]
10. North F, Varkey P, Bartel GA, Cox DL, Jensen PL, Stroebel RJ. Can an office practice telephonic response meet the needs of a pandemic? Telemed J E Health 2010 Dec;16(10):1012-1016. [doi: 10.1089/tmj.2010.0102] [Medline: 21058892]
11. Rubin GJ, Amlôt R, Carter H, Large S, Wessely S, Page L. Reassuring and managing patients with concerns about swine flu: qualitative interviews with callers to NHS Direct. BMC Public Health 2010 Aug 02;10(1):451 [FREE Full text] [doi: 10.1186/1471-2458-10-451] [Medline: 20678192]
12. Cavey AMJ, Spector JM, Ehrhardt D, Kittle T, McNeill M, Greenough PG, et al. Mississippi's infectious disease hotline: a surveillance and education model for future disasters. Prehosp Disaster Med 2009 Jun 28;24(1):11-17. [doi: 10.1017/s1049023x00006488] [Medline: 19557953]
13. North F, Varkey P, Bartel GA, Cox DL, Jensen PL, Stroebel RJ. Can an office practice telephonic response meet the needs of a pandemic? Telemed J E Health 2010 Dec;16(10):1012-1016. [doi: 10.1089/tmj.2010.0102] [Medline: 21058892]
14. Kirkpatrick JN, Hull SC, Fedson S, Mullen B, Goodlin SJ. Scarce-Resource Allocation and Patient Triage During the COVID-19 Pandemic: JACC Review Topic of the Week. J Am Coll Cardiol 2020 Jul 07;76(1):85-92 [FREE Full text] [doi: 10.1016/j.jacc.2020.05.006] [Medline: 32407772]
15. Swiss Academy Of Medical Sciences. COVID-19 pandemic: triage for intensive-care treatment under resource scarcity (3rd, updated version). Swiss Med Wkly 2020 Nov 02;150:w20401 [FREE Full text] [doi: 10.4414/smw.2020.20401] [Medline: 33180951]
16. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, ACCP Task Force for Mass Critical Care. Triage of Scarce Critical Care Resources in COVID-19 An Implementation Guide for Regional Allocation: An Expert Panel Report of the Task Force for Mass Critical Care and the American College of Chest Physicians. Chest 2020 Jul;158(1):212-225 [FREE Full text] [doi: 10.1016/j.chest.2020.03.063] [Medline: 32289312]
17. Coronavirus. Michigan.gov. URL: https://www.michigan.gov/coronavirus [accessed 2021-10-06]
18. Kumar N, Pandey A, Garg N, Sampene E, Lavie CJ. Seasonal and Geographic Patterns in Seeking Cardiovascular Health Information: An Analysis of the Online Search Trends. Mayo Clin Proc 2018 Sep;93(9):1185-1190 [FREE Full text] [doi: 10.1016/j.mayocp.2018.07.011] [Medline: 30193672]
19. Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of Search Engine Queries for Chest Pain With Coronary Heart Disease Epidemiology. JAMA Cardiol 2018 Dec 01;3(12):1218-1221 [FREE Full text] [doi: 10.1001/jamacardio.2018.3459] [Medline: 30422176]
20. Reardon P. Coronavirus hotline supporting South Australians in need. The Advertiser. 2020 May 04. URL: https://www.pressreader.com/australia/the-advertiser/20200504/281771336354391 [accessed 2021-09-27]
21. Travel Leaders Corporate Keeps Travel Advisors On The Payroll By Supporting Call Center Operation for Pandemic Response. PR Newswire. 2020 Apr 30. URL: https://www.prnewswire.com/news-releases/travel-leaders-corporate-keeps-travel-advisors-on-the-payroll-by-supporting-call-center-operation-for-pandemic-response-301050018.html [accessed 2021-09-27]
22. Hogenbirk JC, Pong RW, Lemieux SK. Impact of telephone triage on medical service use: implications for rural and remote areas. J Agric Saf Health 2005 May;11(2):229-237. [doi: 10.13031/2013.18190] [Medline: 15931949]
23. Call centre staff pandemic survey reveals most believe they are non-essential. University of Strathclyde Glasgow. 2020 Apr 17. URL: https://www.strath.ac.uk/whystrathclyde/news/surveyoncallcentrestaffworkingthroughpandemicrevealsmostbelievetheyarenon-essential/ [accessed 2021-09-27]
24. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

25. Stephenson B. ResQ praised for reaction to crisis as hundreds of call centre staff set up at home. Hull Daily Mail. URL: https://www.hulldailymail.co.uk/news/hull-east-yorkshire-news/resq-coronavirus-call-centre-staff-3995262 [accessed 2020-09-21]

Edited by C Basch; submitted 21.02.21; peer-reviewed by A Ren, E van der Velde; comments to author 30.04.21; revised version received 08.06.21; accepted 24.06.21; published 01.11.21.

Please cite as:
Cher BAY, Wilson EA, Pinsky AM, Townshend RF, Wolski AV, Broderick M, Milen AM, Lau A, Singh A, Cinti SK, Engelke CG, Saha AK
Utility of a Telephone Triage Hotline in Response to the COVID-19 Pandemic: Longitudinal Observational Study
J Med Internet Res 2021;23(11):e28105
URL: https://www.jmir.org/2021/11/e28105
doi:10.2196/28105
PMID:34559669

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org